16/PRTS

10/510021

DT04 Rec'd PCT/PTO 01 OCT 2004

Identification of virulence associated regions RD1 and RD5 leading to improve vaccine of *M. bovis* BCG and *M. microti*

Virulence associated regions have been sought for a long time in *Mycobacterium*. The present invention concerns the identification of 2 genomic regions which are shown to be associated with a virulent phenotype in *Mycobacteria* and particularly in *M. tuberculosis*. It concerns also the fragments of said regions.

One of these two regions are known as RD5 as disclosed in Molecular Microbiology (1999), vol. 32, pages 643 to 655 (Gordon S.V. et al.). The other region named RD1-2F9 spans the known region RD1 as disclosed in Molecular Microbiology (1999), vol. 32, pages 643 to 655 (Gordon S.V. et al.). Both of the regions RD1 and RD5 or at least one of them are absent from the vaccine strains of *M. bovis* BCG and in *M. microti*, strains found involved and used as live vaccines in the 1960's.

Other applications which are encompassed by the present invention are related to the use of all or part of the said regions to detect virulent strains of *Mycobacteria* and particularly *M. tuberculosis* in humans and animals. The region RD1-2F9 and RD5 are considered as virulence markers under the present invention.

The recombinant Mycobacteria and particularly *M. bovis* BCG after modification of their genome by introduction of all or part of RD1-2F9 region and/or RD5 region in said genome can be used for the immune system of patients affected with a cancer as for example a bladder cancer.

The present invention relates to a strain of *M. bovis* BCG or *M. microti*, wherein said strain has integrated all or part of the region RD1-2F9 responsible for enhanced immunogenicity to the tubercle bacilli, especially the genes encoding the ESAT-6 and CFP-10 antigenes. These strains will be referred to as the *M. bovis* BCG::RD1 or *M. microti*::RD1 strains and are useful as a new improved vaccine for prevention of tuberculosis infections and for treating superficial bladder cancer.

Mycobacterium bovis BCG (bacille Calmette-Guérin) has been used since 1921 to prevent tuberculosis although it is of limited efficacy against adult pulmonary disease in highly endemic areas. *Mycobacterium microti*, another member of the *Mycobacterium tuberculosis* complex, was originally described as the infective agent of a tuberculosis-like disease in voles (*Microtus agrestis*) in the 1930's (Wells, A. Q. 1937. Tuberculosis in wild voles. Lancet 1221 and Wells, A. Q. 1946. The murine type of tubercle bacillus. Medical Research council special report series 259:1-42.). Until recently, *M. microti* strains were thought to be pathogenic only for voles, but not for humans and some were even used as a live-vaccine. In fact, the vole bacillus proved to be safe and effective in preventing clinical tuberculosis in a trial involving roughly 10,000 adolescents in the UK in the 1950's (Hart, P. D. a., and I. Sutherland. 1977. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. British Medical Journal 2:293-295). At about the same time, another strain, OV166, was successfully administered to half a million newborns in Prague, former Czechoslovakia, without any

serious complications (Sula, L., and I. Radkovsky. 1976. Protective effects of M. microti vaccine against tuberculosis. J. Hyg. Epid. Microbiol. Immunol. 20:1-6). *M. microti* vaccination has since been discontinued because it was no more effective than the frequently employed BCG vaccine. As a result, improved vaccines are needed for preventing and treating tuberculosis.

The problem for attempting to ameliorate this live vaccine is that the molecular mechanism of both the attenuation and the immunogenicity of BCG is still poorly understood. Comparative genomic studies of all six members of the *M. tuberculosis* complex have identified more than 140 genes, whose presence is facultative, that may confer differences in phenotype, host range and virulence. Relative to the genome of the paradigm strain, *M. tuberculosis* H37Rv *(S. T. Cole, et al., Nature 393, 537 (1998))*, many of these genes occur in chromosomal regions that have been deleted from certain species (RD1-16, RvD1-5), M. A. Behr, et al., Science 284, 1520 (1999) ; R. Brosch, et al., Infection Immun. 66, 2221 (1998) ; S. V. Gordon, et al., Molec Microbiol 32, 643 (1999) ; H. Salamon, et al, Genome Res 10, 2044 (2000), G. G. Mahairas et al, J. Bacteriol. 178, 1274 (1996) and R. Brosch, et al., Proc Natl Acad Sci USA 99, 3684 (2002).

In connection with the invention and based on their distribution among tubercle bacilli and potential to encode virulence functions, RD1, RD3-5, RD7 and RD9 (Fig. 1A, B) were accorded highest priority for functional genomic analysis using "knock-ins" of *M. bovis* BCG to assess their potential contribution to the attenuation process. Clones spanning these RD regions were selected from an ordered *M. tuberculosis* H37Rv library of integrating shuttle cosmids (S. T. Cole, *et al.*, *Nature* **393**, 537 (1998) and W. R. Bange, et al, *Tuber. Lung Dis.* **79**, 171 (1999)), and individually electroporated into BCG Pasteur, where they inserted stably into the *attB* site (M. H. Lee, et al, *Proc. Natl. Acad. Sci. USA* **88**, 3111 (1991)).

We have uncovered that only reintroduction of all or part of RD1-2F9 led to profound phenotypic alteration. Strikingly, the BCG::RD1 "knock-in" grew more vigorously than BCG controls in immuno-deficient mice, inducing extensive splenomegaly and granuloma formation.

RD1 is restricted to the avirulent strains *M. bovis* BCG and *M. microti*. Although the endpoints are not identical, the deletions have removed from both vaccine strains a cluster of six genes (Rv3871-Rv3876) that are part of the ESAT-6 locus (Fig. 1A (S. T. Cole, *et al.*, *Nature* **393**, 537 (1998) and F. Tekaia, *et al.*, *Tubercle Lung Disease* **79**, 329 (1999)).

Among the missing products are members of the mycobacterial PE (Rv3872), PPE (Rv3873), and ESAT-6 (Rv3874, Rv3875) protein families. Despite lacking obvious secretion signals, ESAT-6 (Rv3875) and the related protein CFP-10 (Rv3874), are abundant components of short-term culture filtrate, acting as immunodominant T-cell antigens that induce potent Th1 responses (F. Tekaia, et al., Tubercle Lung Disease 79, 329 (1999) ; A. L. Sorensen, et al, Infect. Immun. 63, 1710 (1995) and R. Colangelli, et al., Infect. Immun. 68, 990 (2000)).

In summary, we have discovered that the restoration of RD1-2F9 to *M. bovis* BCG leads to increased persistence in immunocompetent mice. The *M. bovis* BCG::RD1 strain induces RD1-specific immune responses of the Th1-type, has enhanced immunogenicity and confers better protection than *M. bovis* BCG alone in animal models of tuberculosis. The *M. bovis* BCG::RD1 vaccine is significantly more virulent than *M. bovis* BCG in immunodeficient mice but considerably less virulent than *M. tuberculosis*.

In addition, we show that *M. microti* lacks a different but overlapping part of the RD1 region (RD1^{mic}) to *M. bovis* BCG and our results indicate that reintroduction of RD1-2F9 confers increased virulence of BCG ::RD1 in immunodeficient mice. The rare

strains of *M. microti* that are associated with human disease contain a region referred to as $RD5^{mic}$ whereas those from voles do not.

M. bovis BCG vaccine could be improved by reintroducing other genes encoding ESAT-6 family members that have been lost, notably, those found in the RD8 and RD5 loci of *M. tuberculosis*. These regions also code for additional T-cell antigens.

M. bovis BCG::RD1 could be improved by reintroducing the RD8 and RD5 loci of *M. tuberculosis*.

M. bovis BCG vaccine could be improved by reintroducing and overexpressing the genes contained in the RD1, RD5 and RD8 regions.

Accordingly, these new strains, showing greater persistence and enhanced immunogenicity, represent an improved vaccine for preventing tuberculosis and treating bladder cancer.

In addition, the greater persistence of these recombinant strains is an advantage for the presentation of other antigens, for instance from HIV in humans and in order to induce protection immune responses. Those improved strains may also be of use in veterinary medicine, for instance in preventing bovine tuberculosis.

Description

Therefore, the present invention is aimed at a strain of *M. bovis* BCG or *M. microti*, wherein said strain has integrated all or part of the RD1-2F9 region as shown in SEQ ID No 1 responsible for enhanced immunogenicity to the tubercle bacilli. These strains will be referred to as the *M. bovis* BCG::RD1 or *M. microti*::RD1 strains.

In connection with the invention, "part or all of the RD1-2F9 region" means that the strain has integrated a portion of DNA originating from *Mycobacterium tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum*, *M. bovis*, *M. canettii*), which comprises at least one, two, three, four, five, or more gene(s) selected from Rv3861 (SEQ ID No 4), Rv3862 (SEQ ID No 5), Rv3863 (SEQ ID No 6), Rv3864 (SEQ ID No 7), Rv3865 (SEQ ID No 8), Rv3866 (SEQ ID No 9), Rv3867 (SEQ ID No 10), Rv3868 (SEQ ID No 11), Rv3869 (SEQ ID No 12), Rv3870 (SEQ ID No 13), Rv3871 (SEQ ID No 14), Rv3872 (SEQ ID No 15, mycobacterial PE), Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6), Rv3876 (SEQ ID No 19), Rv3877 (SEQ ID No 20), Rv3878 (SEQ ID No 21), Rv3879 (SEQ ID No 22), Rv3880 (SEQ ID No 23), Rv3881 (SEQ ID No 24), Rv3882 (SEQ ID No 25), Rv3883 (SEQ ID No 26), Rv3884 (SEQ ID No 27) and Rv3885 (SEQ ID No 28). The expression "a portion of DNA" means also a nucleotide sequence or a nucleic acid or a polynucleotide. The expression "gene" is referred herein as the coding sequence in frame with its natural promoter as well as the coding sequence which has been isolated and framed with an exogenous promoter, for example a promoter capable of directing high level of expression of said coding sequence.

In a specific aspect, the invention relates to a strain of *M. bovis* BCG or *M. microti* wherein said strain has integrated at least one, two, three or more gene(s) selected from Rv3867 (SEQ ID No 10), Rv3868 (SEQ ID No 11), Rv3869 (SEQ ID No 12), Rv3870 (SEQ ID No 13),Rv3871 (SEQ ID No 14), Rv3872 (SEQ ID No 15, mycobacterial PE), Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6), Rv3876 (SEQ ID No 19) and Rv3877 (SEQ ID No 20).

In an another specific aspect, the invention relates to a strain of *M. bovis* BCG or *M. microti* wherein said strain has integrated at least one, two, three or more gene(s) selected from Rv3871 (SEQ ID No 14), Rv3872 (SEQ ID No 15, mycobacterial PE),

Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6) and Rv3876 (SEQ ID No 19).

Preferably, a strain according to the invention is one which has integrated a portion of DNA originating from *Mycobacterium tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum, M. bovis, M. canettii*), which comprises at least four genes selected from Rv3861 (SEQ ID No 4), Rv3862 (SEQ ID No 5), Rv3863 (SEQ ID No 6), Rv3864 (SEQ ID No 7), Rv3865 (SEQ ID No 8), Rv3866 (SEQ ID No 9), Rv3867 (SEQ ID No 10), Rv3868 (SEQ ID No 11), Rv3869 (SEQ ID No 12), Rv3870 (SEQ ID No 13), Rv3871 (SEQ ID No 14), Rv3872 (SEQ ID No 15, mycobacterial PE), Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6), Rv3876 (SEQ ID No 19), Rv3877 (SEQ ID No 20), Rv3878 (SEQ ID No 21), Rv3879 (SEQ ID No 22), Rv3880 (SEQ ID No 23), Rv3881 (SEQ ID No 24), Rv3882 (SEQ ID No 25), Rv3883 (SEQ ID No 26), Rv3884 (SEQ ID No 27) and Rv3885 (SEQ ID No 28)., provided that it comprises Rv3874 (SEQ ID No 17, CFP-10) and/or Rv3875 (SEQ ID No 18, ESAT-6)..

Strains which have integrated a portion of DNA originating from *Mycobacterium tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum, M. bovis, M. canettii*)comprising at least Rv3871 (SEQ ID No 14), Rv3875 (SEQ ID No 18, ESAT-6) and Rv3876 (SEQ ID No 19) or at least Rv3871 (SEQ ID No 14), Rv3875 (SEQ ID No 18, ESAT-6) and Rv3877 (SEQ ID No 20) or at least Rv3871 (SEQ ID No 14), Rv3875 (SEQ ID No 18, ESAT-6), Rv3876 (SEQ ID No 19) and Rv3877 (SEQ ID No 20) are of particular interest.

The above strains according to the invention may further comprise Rv3874 (SEQ ID No 17, CFP-10), Rv3872 (SEQ ID No 15, mycobacterial PE) and/or Rv3873 (SEQ ID No 16, PPE). In addition, it may further comprise at least one, two, three or four gene(s) selected from Rv3861 (SEQ ID No 4), Rv3862 (SEQ ID No 5), Rv3863 (SEQ ID No 6),

Rv3864 (SEQ ID No 7), Rv3865 (SEQ ID No 8), Rv3866 (SEQ ID No 9), Rv3867 (SEQ ID No 10), Rv3868 (SEQ ID No 11), Rv3869 (SEQ ID No 12), Rv3870 (SEQ ID No 13), Rv3878 (SEQ ID No 21), Rv3879 (SEQ ID No 22), Rv3880 (SEQ ID No 23), Rv3881 (SEQ ID No 24), Rv3882 (SEQ ID No 25), Rv3883 (SEQ ID No 26), Rv3884 (SEQ ID No 27) and Rv3885 (SEQ ID No 28).

The invention encompasses strains which have integrated a portion of DNA originating from *Mycobacterium tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum, M. bovis, M. canettii*), which comprises Rv3875 (SEQ ID No 18, ESAT-6) or Rv3874 (SEQ ID No 17, CFP-10) or both Rv3875 (SEQ ID No 18, ESAT-6) and Rv3874 (SEQ ID No 17, CFP-10).

These genes can be mutated (deletion, insertion or base modification) so as to maintain the improved immunogenicity while decreasing the virulence of the strains. Using routine procedure, the man skilled in the art can select the *M. bovis* BCG::RD1 or *M. microti*::RD1 strains, in which a mutated gene has been integrated, showing improved immunogenicity and lower virulence.

We have shown here that introduction of the RD1-2F9 region makes the vaccine strains induce a more effective immune response against a challenge with *M. tuberculosis*. However, this first generation of constructs can be followed by other, more fine-tuned generations of constructs as the complemented BCG::RD1 vaccine strain also showed a more virulent phenotype in severely immuno-compromised (SCID) mice. Therefore, the BCG::RD1 constructs may be modified so as to be applicable as vaccine strains while being safe for immuno-compromised individuals.The term "construct" means an engineered gene unit, usually involving a gene of interest that has been fused to a promoter.

In this perspective, the man skilled in the art can adapt the BCG::RD1 strain by the conception of BCG vaccine strains that only carry parts of the genes coding for ESAT-6 or CFP-10 in a mycobacterial expression vector (for example pSM81) under the control of a promoter, more particularly an hsp60 promoter. For example, at least one portion of the esat-6 gene that codes for immunogenic 20-mer peptides of ESAT-6 active as T-cell epitopes (Mustafa AS, Oftung F, Amoudy HA, Madi NM, Abal AT, Shaban F, Rosen Krands I, & Andersen P. (2000) Multiple epitopes from the *Mycobacterium tuberculosis* ESAT-6 antigen are recognized by antigen-specific human T cell lines. Clin Infect Dis. 30 Suppl 3:S201-5, peptides P1 to P8 are incorporated herein in the description) could be cloned into this vector and electroporated into BCG, resulting in a BCG strain that produces these epitopes.

Alternatively, the ESAT-6 and CFP-10 encoding genes (for example on plasmid RD1-AP34 and or RD1-2F9) could be altered by directed mutagenesis (using for example QuikChange Site-Directed Mutagenesis Kit from Stratagen) in a way that most of the immunogenic peptides of ESAT-6 remain intact, but the biological functionality of ESAT-6 is lost.

This approach could result in a more protective BCG vaccine without increasing the virulence of the recombinant BCG strain.

Therefore, the invention is also aimed at a method for preparing and selecting *M. bovis* BCG or *M. microti* recombinant strains comprising a step consisting of modifying the *M. bovis* BCG::RD1 or *M. microti*::RD1 strains as defined above by insertion, deletion or mutation in the integrated RD1 region, more particularly in the esat-6 or CFP-10 gene, said method leading to strains that are less virulent for immuno-depressed individuals.

Together, these methods would allow to explain what causes the effect that we see with our BCG::RD1 strain (the presence of additional T-cell epitopes from ESAT-6 and CFP10 resulting in increased immunogenicity) or whether the effect is caused by better

fitness of the recombinant BCG::RD1 clones resulting in longer exposure time of the immune system to the vaccine - or - by a combinatorial effect of both factors.

In a preferred embodiment, the invention is aimed at the *M. bovis* BCG::RD1 strains, which have integrated a cosmid herein referred to as the RD1-2F9 and RD1-AP34 contained in the *E. coli* strains deposited on April 2, 2002 at the CNCM (Institut Pasteur, 25, rue du Docteur Roux, 75724 Paris cedex 15, France) under the accession number I-2831 and I-2832 respectively. The RD1-2F9 is a cosmid comprising the portion of the *Mycobacterium tuberculosis* H37Rv genome previously named RD1-2F9 that spans the RD1 region and contains a gene conferring resistance to Kanamycin. The RD1-AP34 is a cosmid comprising a portion of the *Mycobacterium tuberculosis* H37Rv genome containing two genes coding for ESAT-6 and CFP-10 as well as a gene conferring resistance to Kanamycin.

The cosmid RD1-AP34 contains a 3909 bp fragment of the *M. tuberculosis* H37Rv genome from region 4350459 bp to 4354367 bp that has been cloned into an integrating vector pKint in order to be integrated in the genome of *Mycobacterium bovis* BCG and *Mycobacterium microti* strains (SEQ ID No 3). The Accession No. of the segment 160 of the *M. tuberculosis* H37Rv genome that contains this region is AL022120.

SEQ ID No 3 :

1 - **gaattc**ccat ccagtgagtt caaggtcaag cggcgccccc ctggccaggc atttctcgtc
61 - tcgccagacg gcaaagaggt catccaggcc ccctacatcg agcctccaga agaagtgttc
121 - gcagcacccc caagcgccgg ttaagattat ttcattgccg gtgtagcagg acccgagctc
181 - agcccggtaa tcgagttcgg gcaatgctga ccatcgggtt tgtttccggc tataaccgaa
241 - cggtttgtgt acgggataca aatacaggga gggaagaagt aggcaa*atgg aaaaaatgtc*

301 - acatgatccg atcgctgccg acattggcac gcaagtgagc gacaacgctc tgcacggcgt

361 - gacggccggc tcgacggcgc tgacgtcggt gaccgggctg gttcccgcgg gggccgatga

421 - ggtctccgcc caagcggcga cggcgttcac atcggagggc atccaattgc tggcttccaa

481 - tgcatcggcc caagaccagc tccaccgtgc gggcgaagcg gtccaggacg tcgcccgcac

*541 - ctattcgcaa atcgacgacg gcgccgccgg cgtcttcgcc gaa*taggccc ccaacacatc

601 - ggagggagtg atcaccatgc tgtggcacgc aatgccaccg gagctaaata ccgcacggct

661 - gatggccggc gcgggtccgg ctccaatgct tgcggcggcc gcgggatggc agacgctttc

721 - ggcggctctg gacgctcagg ccgtcgagtt gaccgcgcgc ctgaactctc tgggagaagc

781 - ctggactgga ggtggcagcg acaaggcgct tgcggctgca acgccgatgg tggtctggct

841 - acaaaccgcg tcaacacagg ccaagacccg tgcgatgcag gcgacggcgc aagccgcggc

901 - atacacccag gccatggcca cgacgccgtc gctgccggag atcgccgcca accacatcac

961 - ccaggccgtc cttacggcca ccaacttctt cggtatcaac acgatcccga tcgcgttgac

1021 - cgagatggat tatttcatcc gtatgtggaa ccaggcagcc ctggcaatgg aggtctacca

1081 - ggccgagacc gcggttaaca cgcttttcga gaagctcgag ccgatggcgt cgatccttga

1141 - tcccggcgcg agccagagca cgacgaaccc gatcttcgga atgccctccc ctggcagctc

1201 - aacaccggtt ggccagttgc cgccggcggc tacccagacc ctcggccaac tgggtgagat

1261 - gagcggcccg atgcagcagc tgacccagcc gctgcagcag gtgacgtcgt tgttcagcca

1321 - ggtgggcggc accggcggcg gcaacccagc cgacgaggaa gccgcgcaga tgggcctgct

1381 - cggcaccagt ccgctgtcga accatccgct ggctggtgga tcaggcccca gcgcgggcgc

1441 - gggcctgctg cgcgcggagt cgctacctgg cgcaggtggg tcgttgaccc gcacgccgct

1501 - gatgtctcag ctgatcgaaa agccggttgc cccctcggtg atgccggcgg ctgctgccgg

1561 - atcgtcggcg acgggtggcg ccgctccggt gggtgcggga gcgatgggcc agggtgcgca

1621 - atccggcggc tccaccaggc cgggtctggt cgcgccggca ccgctcgcgc aggagcgtga

1681 - agaagacgac gaggacgact gggacgaaga ggacgactgg tgagctcccg taatgacaac

1741 - agacttcccg gccacccggg ccggaagact tgccaacatt ttggcgagga aggtaaagag

1801 - agaaagtagt ccagcatggc **agagatgaag accgatgccg ctaccctcgc gcaggaggca**

1861 - ggtaatttcg agcggatctc cggcgacctg aaaacccaga tcgaccaggt ggagtcgacg

1921 - gcaggttcgt tgcagggcca gtggcgcggc gcggcgggga cggccgccca ggccgcggtg

1981 - **gtgcgcttcc aagaagcagc caataagcag aagcaggaac tcgacgagat ctcgacgaat**

2041 - **attcgtcagg ccggcgtcca atactcgagg gccgacgagg agcagcagca ggcgctgtcc**

2101 - **tcgcaaatgg gctt**ctgacc cgctaatacg aaaagaaacg gagcaaaaac *atgacagagc*

2161 - *agcagtggaa tttcgcgggt atcgaggccg cggcaagcgc aatccaggga aatgtcacgt*

2221 - *ccattcattc cctccttgac gaggggaagc agtccctgac caagctcgca gcggcctggg*

2281 - *gcggtagcgg ttcggaggcg taccagggtg tccagcaaaa atgggacgcc acggctaccg*

2341 - *agctgaacaa cgcgctgcag aacctggcgc ggacgatcag cgaagccggt caggcaatgg*

2401 - *cttcgaccga aggcaacgtc actgggatgt tcgca*taggg caacgccgag ttcgcgtaga

2461 - atagcgaaac acgggatcgg gcgagttcga ccttccgtcg gtctcgccct ttctcgtgtt

2521 - tatacgtttg agcgcactct gagaggttgt catggcggcc gactacgaca agctcttccg

2581 - gccgcacgaa ggtatggaag ctccggacga tatggcagcg cagccgttct tcgaccccag

2641 - tgcttcgttt ccgccggcgc ccgcatcggc aaacctaccg aagcccaacg gccagactcc

2701 - gcccccgacg tccgacgacc tgtcggagcg gttcgtgtcg gccccgccgc cgccaccccc

2761 - acccccacct ccgcctccgc caactccgat gccgatcgcc gcaggagagc cgccctcgcc

2821 - ggaaccggcc gcatctaaac cacccacacc ccccatgccc atcgccggac ccgaaccggc

2881 - cccacccaaa ccacccacac cccccatgcc catcgccgga cccgaaccgg ccccacccaa

2941 - accacccaca cctccgatgc ccatcgccgg acctgcaccc accccaaccg aatcccagtt

3001 - ggcgcccccc agaccaccga caccacaaac gccaaccgga gcgccgcagc aaccggaatc

3061 - accggcgccc cacgtaccct cgcacgggcc acatcaaccc cggcgcaccg caccagcacc

3121 - gccctgggca aagatgccaa tcggcgaacc cccgcccgct ccgtccagac cgtctgcgtc

3181 - cccggccgaa ccaccgaccc ggcctgcccc ccaacactcc cgacgtgcgc gccggggtca

3241 - ccgctatcgc acagacaccg aacgaaacgt cgggaaggta gcaactggtc catccatcca

3301 - ggcgcggctg cgggcagagg aagcatccgg cgcgcagctc gcccccggaa cggagccctc

3361 - gccagcgccg ttgggccaac cgagatcgta tctggctccg cccacccgcc ccgcgccgac

3421 - agaacctccc cccagcccct cgccgcagcg caactccggt cggcgtgccg agcgacgcgt

3481 - ccaccccgat ttagccgccc aacatgccgc ggcgcaacct gattcaatta cggccgcaac

3541 - cactggcggt cgtcgccgca agcgtgcagc gccggatctc gacgcgacac agaaatcctt

3601 - aaggccggcg gccaaggggc cgaaggtgaa gaaggtgaag ccccagaaac cgaaggccac

3661 - gaagccgccc aaagtggtgt cgcagcgcgg ctggcgacat tgggtgcatg cgttgacgcg

3721 - aatcaacctg ggcctgtcac ccgacgagaa gtacgagctg gacctgcacg ctcgagtccg

3781 - ccgcaatccc cgcgggtcgt atcagatcgc cgtcgtcggt ctcaaaggtg gggctggcaa

3841 - aaccacgctg acagcagcgt tggggtcgac gttggctcag gtgcgggccg accggatcct

3901 - ggc**tctaga**

pos. 0001-0006 **EcoRI-restriction site**

pos. 0286-0583 *Rv3872 coding for a PE-Protein* (SEQ ID No 15)

pos. 0616-1720 Rv3873 coding for a PPE-Protein (SEQ ID No 16)

pos. 1816-2115 **Rv3874 coding for Culture Filtrat protein 10kD (CFP10)** (SEQ ID No 17)

pos. 2151-2435 *Rv3875 coding for Early Secreted Antigen Target 6kD (ESAT6)* (SEQ ID No 18)

pos. 3903-3609 **XbaI-restriction site**

pos. 1816-2435 CFP-10 gene + esat-6 gene (SEQ ID No 29).

These sequences can be completed with the Rv3861 to Rv3871, and Rv3876 to Rv3885 as referred in **Table 1** below.

Gene Name	Gene length	Protein length	Gene type	Accession number in NCBI Bank NC = gene NP = protein	Loc (kb) in M. tuberculosis H37Rv	Coordinates in Mycobacterium tuberculosis H37Rv	Molecular mass of protein (Dalton)	Description
Rv3861	324	108	CDS		4337.95	4337946 .. 4338269	11643.42	hypothetical protein
Rv3862 c-whiB6	348	116	CDS		4338.52	compl 4338174.. 4338521	12792.38	possible transcriptional regulatory protein whiB-like WhiB6
Rv3863	1176	392	CDS		4338.85	4338849.. 4340024	41087.44	hypothetical alanine rich protein
Rv3864	1206	402	CDS		4340.27	4340270..	42068.66	conserved

						4341475		hypothetical protein
Rv3865	309	103	CDS		4341.57	4341566.. 4341874	10618.01	conserved hypothetical protein
Rv3866	849	283	CDS		4341.88	4341880.. 4342728	30064.04	conserved hypothetical protein
Rv3867	549	183	CDS	NC_000962 NP_218384	4342.77	4342767 .. 4343318	19945.52	conserved protein
Rv3868	1719	573	CDS	NC_000962 NP_218385	4343.3	4343311 .. 4345032	62425.40	conserved protein
Rv3869	1440	480	CDS	NC_000962 NP_218386	4345.04	4345036 .. 4346478	51092.58	possible conserved membrane protein
Rv3870	2241	747	CDS	NC_000962 NP_218387	4346.48	4346478 .. 4348721	80912.76	possible conserved membrane protein
Rv3871	1773	591	CDS	NC_000962 NP_218388	4348.83	4348824 .. 4350599	64560.65	conserved protein
Rv3876	1998	666	CDS	NC_000962 NP_218393	4353.01	4353007 .. 4355007	70644.92	conserved proline and alanine rich protein
Rv3877	1533	511	CDS	NC_000962 NP_218394	4355.01	4355004 .. 4356539	53981.12	probable conserved transmembrane protein
Rv3878	840	280	CDS	NC_000962	4356.69	4356693.. 4357532	27395.23	conserved hypothetical alanine rich protein
Rv3879 c	2187	729	CDS	NC_000962	4359.78	compl. 4357596.. 4359782	74492.13	hypothetical alanine and proline rich protein
Rv3880 c	345	115	CDS	NC_000962	4360.55	compl. 4360202.. 4360546	12167.51	conserved hypothetical protein
Rv3881 c	1380	460	CDS	NC_000962	4361.92	compl. 4360546.. 4361925	47593.62	conserved hypothetical alanine and glycine rich protein
Rv3882 c	1386	462	CDS	NC_000962	4363.42	compl. 4362035..	50396.58	possible conser ved membrane

						4363420		protein
Rv3883 c	1338	446	CDS	NC_000962	4364.76	compl. 4363420.: 4364757	45085.89	possible secreted protease
Rv3884 c	1857	619	CDS	NC_000962	4366.84	compl. 4364982.. 4366838	68040.97	probable CBXX/CFQX family protein
Rv3885 c	1611	537	CDS	NC_000962	4368.52	compl. 4366911.. 4368521	57637.95	possible conserved membrane protein

The sequence of the fragment RD1-2F9 (~ 32 kb) covers the region of *the M. tuberculosis* genome AL123456 from ca 4337 kb to ca. 4369 kb, and also contains the sequence described in SEQ ID No 1. Therefore, the invention also embraces *M. bovis* BCG::RD1 strain and *M. microti*::RD1 strain which have integrated the sequence as shown in SEQ ID No 1.

The above described strains fulfill the aim of the invention which is to provide an improved tuberculosis vaccine or *M. bovis* BCG-based prophylactic or therapeutic agent, or a recombinant *M. microti* derivative for these purposes.

The above described *M. bovis* BCG::RD1 strains are better tuberculosis vaccines than *M. bovis* BCG. These strains can also be improved by reintroducing other genes found in the RD8 and RD5 loci of *M. tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum*, *M. bovis*, *M. canettii*). These regions code for additional T-cell antigens.
As indicated, overexpressing the genes contained in the RD1, RD5 and RD8 regions by means of exogenous promoters is encompassed by the invention. The same applies regarding *M. microti*::RD1 strains. *M. microti* strains could also be improved by

reintroducing the RD8 locus of *M. tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum*, *M. bovis*, *M. canettii*).

In a second embodiment, the invention is directed to a cosmid or a plasmid, more commonly named vectors, comprising all or part of the RD1-2F9 region originating from *Mycobacterium tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum*, *M. bovis*, *M. canettii*), said region comprising at least one, two, three or more gene(s) selected from Rv3861 (SEQ ID No 4), Rv3862 (SEQ ID No 5), Rv3863 (SEQ ID No 6), Rv3864 (SEQ ID No 7), Rv3865 (SEQ ID No 8), Rv3866 (SEQ ID No 9), Rv3867 (SEQ ID No 10), Rv3868 (SEQ ID No 11), Rv3869 (SEQ ID No 12), Rv3870 (SEQ ID No 13), Rv3871 (SEQ ID No 14), Rv3872 (SEQ ID No 15, mycobacterial PE), Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6), Rv3876 (SEQ ID No 19), Rv3877 (SEQ ID No 20), Rv3878 (SEQ ID No 21), Rv3879 (SEQ ID No 22), Rv3880 (SEQ ID No 23), Rv3881 (SEQ ID No 24), Rv3882 (SEQ ID No 25), Rv3883 (SEQ ID No 26), Rv3884 (SEQ ID No 27) and Rv3885 (SEQ ID No 28). The term "vector" refers to a DNA molecule originating from a virus, a bacteria, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vectors capacity for self-replication; a vector introduces foreign DNA into host cells, where it can be reproduced in large quantities. Examples are plasmids, cosmids, and yeast artificial chromosomes; vectors are often recombinant molecules containing DNA sequences from several sources.

Preferably, a cosmid or a plasmid of the invention comprises a part of the RD1-2F9 region originating from *Mycobacterium tuberculosis* or any virulent member of the

Mycobacterium tuberculosis complex (*M. africanum, M. bovis, M. canettii*), said part comprising at least one, two, three or more gene(s) selected from Rv3867 (SEQ ID No 10), Rv3868 (SEQ ID No 11), Rv3869 (SEQ ID No 12), Rv3870 (SEQ ID No 13), Rv3871 (SEQ ID No 14), Rv3872 (SEQ ID No 15, mycobacterial PE), Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6), Rv3876 (SEQ ID No 19) and Rv3877 (SEQ ID No 20).

Preferably, a cosmid or a plasmid of the invention comprises a part of the RD1-2F9 region originating from *Mycobacterium tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum, M. bovis, M. canettii*), said part comprising at least one, two, three or more gene(s) selected from Rv3872 (SEQ ID No 15, mycobacterial PE), Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6) and Rv3876 (SEQ ID No 19).

Preferably, a cosmid or a plasmid of the invention comprises CFP-10, ESAT-6 or both or a part of them. It may also comprise a mutated gene selected CFP-10, ESAT-6 or both, said mutated gene being responsible for the improved immunogenicity and decreased virulence.

A cosmid or a plasmid as mentioned above may comprise at least four genes selected from Rv3861 (SEQ ID No 4), Rv3862 (SEQ ID No 5), Rv3863 (SEQ ID No 6), Rv3864 (SEQ ID No 7), Rv3865 (SEQ ID No 8), Rv3866 (SEQ ID No 9), Rv3867 (SEQ ID No 10), Rv3868 (SEQ ID No 11), Rv3869 (SEQ ID No 12), Rv3870 (SEQ ID No 13), Rv3871 (SEQ ID No 14), Rv3872 (SEQ ID No 15, mycobacterial PE), Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6), Rv3876 (SEQ ID No 19), Rv3877 (SEQ ID No 20), Rv3878 (SEQ ID No 21), Rv3879 (SEQ ID No 22), Rv3880 (SEQ ID No 23), Rv3881 (SEQ ID No 24), Rv3882 (SEQ ID No 25), Rv3883 (SEQ ID No 26), Rv3884 (SEQ ID No 27) and Rv3885 (SEQ ID No

28), provided that it comprises Rv3874 (SEQ ID No 17, CFP-10) and/or Rv3875 (SEQ ID No 18, ESAT-6)

Advantageously, a cosmid or a plasmid of the invention comprises a portion of DNA originating from *Mycobacterium tuberculosis* or any virulent member of the *Mycobacterium tuberculosis* complex (*M. africanum*, *M. bovis*, *M. canettii*), which comprises at least Rv3871 (SEQ ID No 14), Rv3875 (SEQ ID No 18, ESAT-6) and Rv3876 (SEQ ID No 19) or at least Rv3871 (SEQ ID No 14), Rv3875 (SEQ ID No 18, ESAT-6) and Rv3877 (SEQ ID No 20) or at least Rv3871 (SEQ ID No 14), Rv3875 (SEQ ID No 18, ESAT-6), Rv3876 (SEQ ID No 19) and Rv3877 (SEQ ID No 20).

The above cosmids or plasmids may further comprise Rv3872 (SEQ ID No 15, mycobacterial PE) Rv3873 (SEQ ID No 16, PPE) Rv3874 (SEQ ID No 17, CFP-10). It may also further comprise at least one, two, three or four gene(s) selected from Rv3861 (SEQ ID No 4), Rv3862 (SEQ ID No 5), Rv3863 (SEQ ID No 6), Rv3864 (SEQ ID No 7), Rv3865 (SEQ ID No 8), Rv3866 (SEQ ID No 9), Rv3867 (SEQ ID No 10), Rv3868 (SEQ ID No 11), Rv3869 (SEQ ID No 12), Rv3870 (SEQ ID No 13), Rv3878 (SEQ ID No 21), Rv3879 (SEQ ID No 22), Rv3880 (SEQ ID No 23), Rv3881 (SEQ ID No 24), Rv3882 (SEQ ID No 25), Rv3883 (SEQ ID No 26), Rv3884 (SEQ ID No 27) and Rv3885 (SEQ ID No 28).

Two particular cosmids of the invention are the cosmids herein referred as RD1-2F9 and RD1-AP34 contained in the *E. coli* strains deposited at the CNCM (Institut Pasteur, 25, rue du Docteur Roux, 75724 Paris cedex 15, France) under the accession number I-2831 and I-2832 respectively.

A particular plasmid or cosmid of the invention is one which has integrated the complete RD1-2F9 region as shown in SEQ ID No 1.

The invention also relates to the use of these cosmids or plasmids for transforming *M. bovis* BCG or *M. microti* strains.

As indicated above, these cosmids or plasmids may comprise a mutated gene selected from Rv3861 to Rv3885, said mutated gene being responsible for the improved immunogenicity and decreased virulence.

In another embodiment, the invention embraces a pharmaceutical composition comprising a strain as depicted above and a pharmaceutically acceptable carrier.

In addition to the strains, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the living vaccine into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.).

Preferably, such composition is suitable for oral, intravenous or subcutaneous administration.

The determination of the effective dose is well within the capability of those skilled in the art. A therapeutically effective dose refers to that amount of active ingredient, i.e the number of strains administered, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices

are preferred. Of course, ED50 is to be modulated according to the mammal to be treated or vaccinated. In this regard, the invention contemplates a composition suitable for human administration as well as veterinary composition.

The invention is also aimed at a vaccine comprising a *M. bovis* BCG::RD1 or *M. microti*::RD1 strain as depicted above and a suitable carrier. This vaccine is especially useful for preventing tuberculosis. It can also be used for treating bladder cancer.

The *M. bovis* BCG::RD1 or *M. microti*::RD1 strains are also useful as a carrier for the expression and presentation of foreign antigens or molecules of interest that are of therapeutic or prophylactic interest. Owing to its greater persistence, BCG::RD1 will present antigens to the immune system over a longer period thereby inducing stronger, more robust immune responses and notably protective responses. Examples of such foreign antigens can be found in patents and patent applications US 6,191,270 for antigen LSA3, US 6,096,879 and US 5,314,808 for HBV antigens, EP 201,540 for HIV-1 antigens, US 5,986,051 for *H. pylori* antigens and FR 2,744,724 for *P. falciparum* MSP-1 antigen.

The invention also concerns a product comprising a strain as depicted above and at least one protein selected from ESAT-6 and CFP-10 or epitope derived thereof for a separate, simultaneous or sequential use for treating tuberculosis.

In still another embodiment, the invention concerns the use of a *M. bovis* BCG::RD1 or *M. microti*::RD1 strain as depicted above for preventing or treating tuberculosis.
It also concerns the use of a *M. bovis* BCG::RD1 or *M. microti*::RD1 strain as a powerful adjuvant/immunomodulator used in the treatment of superficial bladder cancer.

The invention also contemplates the identification at the species level of members of the *M. tuberculosis* complex by means of an RD-based molecular diagnostic test. Inclusion

of markers for $RD1^{mic}$ and $RD5^{mic}$ would improve the tests and act as predictors of virulence, especially in humans.

In this regard, the invention concerns a diagnostic kit for the identification at the species level of members of the *M. tuberculosis* complex comprising DNA probes and primers specifically hybridizing to a DNA portion of the RD1 or RD5 region of *M. tuberculosis*, more particularly probes hybridizing under stringent conditions to a gene selected from Rv3871 (SEQ ID No 14), Rv3872 (SEQ ID No 15, mycobacterial PE), Rv3873 (SEQ ID No 16, PPE), Rv3874 (SEQ ID No 17, CFP-10), Rv3875 (SEQ ID No 18, ESAT-6), and Rv3876 (SEQ ID No 19), preferably CFP-10 and ESAT-6.

As used herein, the term "stringent conditions" refers to conditions which permit hybridization between the probe sequences and the polynucleotide sequence to be detected. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

Among the preferred primers, we can cite:

primer esat-6F GTCACGTCCATTCATTCCCT (SEQ ID No 32),

primer esat-6R ATCCCAGTGACGTTGCCTT) (SEQ ID No 33),

primer $RD1^{mic}$ flanking region F GCAGTGCAAAGGTGCAGATA (SEQ ID No 34),

primer $RD1^{mic}$ flanking region R GATTGAGACACTTGCCACGA (SEQ ID No 35),

primer $RD5^{mic}$ flanking region F GAATGCCGACGTCATATCG (SEQ ID No 39),

primer RD5^{mic} flanking region R CGGCCACTGAGTTCGATTAT (SEQ ID No 40).

The present invention covers also the complementary nucleotide sequences of said above primers as well as the nucleotide sequences hybridizing under stringent conditions with them and having at least 20 nucleotides and less than 500 nucleotides.

Diagnostic kits for the identification at the species level of members of the *M. tuberculosis* complex comprising at least one, two, three or more antibodies directed to mycobacterial PE, PPE, CFP-10, ESAT-6, are also embraced by the invention.

Preferably, such kit comprises antibodies directed to CFP-10 and ESAT-6.

As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fab, F(ab').sub.2, and Fv, which are capable of binding the epitopic determinant. Probes or antibodies can be labeled with isotopes, fluorescent or phosphorescent molecules or by any other means known in the art.

The invention also relates to virulence markers associated with RD1 and/or RD5 regions of the genome of *M. tuberculosis* or a part of these regions.

The invention is further detailed below and will be illustrated with the following figures.

Figure legends

Figure 1: *M. bovis* BCG and *M. microti* have a chromosomal deletion, RD1, spanning the *cfp10-esat6* locus.

(A) Map of the *cfp10-esat6* region showing the six possible reading frames and the *M. tuberculosis* H37Rv gene predictions. This map is also available at: (http://genolist.pasteur.fr/TubercuList/).

The deleted regions are shown for BCG and *M. microti* with their respective H37Rv genome coordinates, and the extent of the conserved ESAT-6 locus (F. Tekaia, *et al.*, *Tubercle Lung Disease* 79, 329 (1999)), is indicated by the gray bar.

(B) Table showing characteristics of deleted regions selected for complementation analysis. Potential virulence factors and their putative functions disrupted by each deletion are shown. The coordinates are for the *M. tuberculosis* H37Rv genome.

(C) Clones used to complement BCG. Individual clones spanning RD1 regions (RD1-I106 and RD1-2F9) were selected from an ordered *M. tuberculosis* genomic library (R.B. unpublished) in pYUB412 (S. T. Cole, *et al.*, *Nature* **393**, 537 (1998) and W. R. Bange, F. M. Collins, W. R. Jacobs, Jr., *Tuber. Lung Dis.* **79**, 171 (1999)) and electroporated into *M. bovis* BCG strains, or *M. microti.* Hygromycin-resistant transformants were verified using PCR specific for the corresponding genes. pAP35 was derived from RD1-2F9 by excision of an *Afl*II fragment. pAP34 was constructed by subcloning an *Eco*RI-*Xba*I fragment into the integrative vector pKINT. The ends of each fragment are related to the BCG RD1 deletion (shaded box) with black lines and the H37Rv coordinates for the other fragment ends given in kilobases.

(D) Immunoblot analysis, using an ESAT-6 monoclonal antibody, of whole cell protein extracts from log-phase cultures of (well n°1) H37Rv (S. T. Cole, *et al.*, *Nature* **393**, 537 (1998)), (n°2) BCG::pYUB412 (M. A. Behr, *et al.*, *Science* **284**, 1520 (1999)), (n°3) BCG::RD1-I106 (R. Brosch, *et al.*, *Infection Immun.* **66**, 2221 (1998)), (n°4) BCG::RD1-2F9 (S. V. Gordon, *et al.*, *Molec Microbiol* **32**, 643 (1999)), (n°5) *M. bovis* (H. Salamon et al, *Genome Res* **10**, 2044 (2000)), (n°6) *Mycobacterium smegmatis* (G.

G. Mahairas, et al, *J. Bacteriol.* 178, 1274 (1996)), (n°7) *M. smegmatis*::pYUB412, and (n°8) *M. smegmatis*:: RD1-2F9 (R. Brosch, *et al.*, *Proc Natl Acad Sci USA* 99, 3684 (2002)).

Figure 2: Complementation of BCG Pasteur with the RD1 region alters the colony morphology and leads to accumulation of Rv3873 and ESAT-6 in the cell wall.

(A) Serial dilutions of 3 week old cultures of BCG::pYUB412, BCG::I106 or BCG::RD1-2F9 growing on Middlebrook 7H10 agar plates. The white square shows the area of the plate magnified in the image to the right.

(B) Light microscope image at fifty fold magnification of BCG::pYUB412 and BCG::RD1-2F9 colonies. 5 µl drops of bacterial suspensions of each strain were spotted adjacently onto 7H10 plates and imaged after 10 days growth, illuminating the colonies through the agar.

(C) Immunoblot analysis of different cell fractions of H37Rv obtained from http://www.cvmbs.colostate.edu/microbiology/tb/ResearchMA.html using either an anti-ESAT-6 antibody or

(D) anti-Rv3873 (PPE) rabbit serum. H37Rv and BCG signify whole cell extracts from the respective bacteria and Cyt, Mem and CW correspond to the cytosolic, membrane and cell wall fractions of *M. tuberculosis* H37Rv.

Figure 3: Complementation of BCG Pasteur with the RD1 region increases bacterial persistence and pathogenicity in mice.

(A) Bacteria in the spleen and lungs of BALB/c mice following intravenous (i.v.) infection via the lateral tail vein with 10^6 colony forming units (cfu) of *M. tuberculosis*

H37Rv (black) or 10^7 cfu of either BCG::pYUB412 (light grey) or BCG::RD1-I106 (grey).

(B) Bacterial persistence in the spleen and lungs of C57BL/6 mice following i.v. infection with 10^5 cfu of BCG::pYUB412 (light grey), BCG::RD1-I106 (middle grey) or BCG::RD1-2F9 (dark grey).

(C) Bacterial multiplication after i.v. infection with 10^6 cfu of BCG::pYUB412 (light grey) and BCG::RD1-2F9 (grey) in severe combined immunodeficiency mice (SCID). For **A**, **B**, and **C** each timepoint is the mean of 3 to 4 mice and the error bars represent standard deviations.

(D) Spleens from SCID mice three weeks after i.v. infection with 10^6 cfu of either BCG::pYUB412, BCG::RD1-2F9 or BCG::I301 (an RD3 "knock-in", Fig. 1B). The scale is in cm.

Figure 4: Immunisation of mice with BCG::RD1 generates marked ESAT-6 specific T-cell responses and enhanced protection to a challenge with *M. tuberculosis*.

(A) Proliferative response of splenocytes of C57BL/6 mice immunised subcutaneously (s.c.) with 10^6 CFU of BCG::pYUB412 (open squares) or BCG::RD1-2F9 (solid squares) to *in vitro* stimulation with various concentrations of synthetic peptides from poliovirus type 1 capsid protein VP1, ESAT-6 or Ag85A (K. Huygen, et al., *Infect. Immun.* **62**, 363 (1994), L. Brandt, *J.Immunol.* **157**, 3527 (1996) and C. Leclerc et al, *J. Virol.* **65**, 711 (1991)).

(B) Proliferation of splenocytes from BCG::RD1-2F9-immunised mice in the absence or presence of 10 μg/ml of ESAT-6 1-20 peptide, with or without 1 μg/ml of anti-CD4 (GK1.5) or anti-CD8 (H35-17-2) monoclonal antibody. Results are expressed as mean and standard deviation of ^{3}H-thymidine incorporation from duplicate wells.

(C) Concentration of IFN-γ in culture supernatants of splenocytes of C57BL/6 mice stimulated for 72 h with peptides or PPD after s.c. or i.v. immunisation with either BCG::pYUB412 (middle grey and white) or BCG::RD1-2F9 (light grey and black). Mice were inoculated with either 10^6 (white and light grey) or 10^7(middle grey and black) cfu. Levels of IFN-γ were quantified using a sandwich ELISA (detection limit of 500 pg/ml) with the mAbs R4-6A2 and biotin-conjugated XMG1.2. Results are expressed as the mean and standard deviation of duplicate culture wells.

(D) Bacterial counts in the spleen and lungs of vaccinated and unvaccinated BALB/c mice 2 months after an i.v. challenge with *M. tuberculosis* H37Rv. The mice were challenged 2 months after i.v. inoculation with 10^6 cfu of either BCG::pYUB412 or BCG::RD1-2F9. Organ homogenates for bacterial enumeration were plated on 7H11 medium, with or without hygromycin, to differentiate *M. tuberculosis* from residual BCG colonies. Results are expressed as the mean and standard deviation of 4 to 5 mice and the levels of significance derived using the Wilcoxon rang sum test.

Figure 5: *Mycobacterium microti* strain OV254 BAC map (BAC clones named MiXXX, where XXX is the identification number of the clone), overlaid on the *M. tuberculosis* H37Rv (BAC clones named RvXXX, where XXX is the identification number of the clone) and *M. bovis* AF2122/97 (BAC clones named MbXXX, where XXX is the identification number of the clone) BAC maps. The scale bars indicate the position on the *M. tuberculosis* genome.

Figure 6: Difference in the region 4340-4360 kb between the deletion in BCG $RD1^{bcg}$ (A) and in *M. microti* $RD1^{mic}$ (C) relative. to *M. tuberculosis* H37Rv (B).

Figure 7: Difference in the region 3121-3127 kb between *M. tuberculosis* H37Rv (A) and *M. microti* OV254 (B). Gray boxes picture the direct repeats (DR), black ones the

unique numbered spacer sequences. * spacer sequence identical to the one of spacer 58 reported by van Embden *et al.* (42). Note that spacers 33-36 and 20-22 are not shown because H37Rv lacks these spacers.

Figure 8: A) *Ase*I PFGE profiles of various *M. microti* strains; Hybridization with a radiolabeled **B)** *esat-6* probe; **C)** probe of the RD1^{mic} flanking region; **D)** *plcA* probe. 1. *M. bovis* AF2122/97, 2. *M. canetti*, 3. *M. bovis* BCG Pasteur, 4. *M. tuberculosis* H37Rv, 5. *M. microti* OV254, 6. *M. microti* Myc 94-2272, 7. *M. microti* B3 type mouse, 8. *M. microti* B4 type mouse , 9. *M. microti* B2 type llama, 10. *M. microti* B1 type llama, 11. *M. microti* ATCC 35782. M: Low range PFGE marker (NEB).

Figure 9: PCR products obtained from various *M. microti* strains using primers that flank the RD1^{mic} region, for amplifying ESAT-6 antigen, that flank the MiD2 region. 1. *M. microti* B1 type llama, 2. *M. microti* B4 type mouse, 3. *M. microti* B3 type mouse, 4. *M. microti* B2 type llama, 5. *M. microti* ATCC 35782, 6. *M. microti* OV254, 7. *M. microti* Myc 94-2272, 8. *M. tuberculosis* H37Rv.

Figure 10: Map of the *M. tuberculosis* H37Rv RD1 genomic region. Map of the fragments used to complement BCG and *M. microti* (black) and the genomic regions deleted from different mycobacterial strains (grey). The middle part shows key genes, putative promoters (P) and transcripts, the various proteins from the RD1 region, their sizes (number of amino acid residues), InterPro domains (http://www.ebi.ac.uk/interpro/), membership of *M. tuberculosis* protein families from TubercuList (http://genolist.pasteur.fr/TubercuList/). The dashed lines mark the extent of the RD1 deletion in BCG, *M. microti* and *M. tuberculosis* clinical isolate MT56 (Brosch, R., *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 99, 3684-9. (2002)). *M. bovis* AF2122/97 is shown because it contains a frameshift

mutation in Rv3881, a gene flanking the RD1 region of BCG. The fragments are drawn to show their ends in relation to the genetic map, unless they extend beyond the genomic region indicated. pRD1-2F9, pRD1-I106 and pAP35 are based on pYUB412; pAP34 on pKINT; pAP47 and pAP48 on pSM81.

Figure 11: Western blot analysis of various RD1 knock-ins of *M. bovis* BCG and *M. microti*. The left panel shows results of immunodetection of ESAT-6, CFP-10 and PPE68 (Rv3873) in whole cell lysates (WCL) and culture supernatants of BCG; the centre panel displays the equivalent findings from *M. microti* and the right panel contains *M. tuberculosis* H37Rv control samples. Samples from mycobacteria transformed with the following plasmids were present in lanes: -, pYUB412 vector control; 1, pAP34; 2, pAP35; 3, RD1-I106; 4, RD1-2F9. The positions of the nearest molecular weight markers are indicated.

Figure 12: Analysis of immune responses induced by BCG recombinants. **A,** The upper three panels display the results of splenocyte proliferation assays in response to stimulation *in vitro* with a peptide from MalE (negative control), to PPD or to a peptide containing an immunodominant CD4-epitope from ESAT-6. **B,** The lower panel shows IFN-γ production by splenocytes in response to the same antigens. Symbols indicate the nature of the various BCG transformants. Samples were taken from C57BL/6 mice immunised subcutaneously.

Figure 13: Further immunological characterization of responses to BCG::RD1-2F9 **A,** Proliferative response of splenocytes of C57BL/6 mice immunised subcutaneously (s.c.) with 10^6 CFU of BCG::pYUB412 or BCG::RD1-2F9 to *in vitro* stimulation with various concentrations of synthetic peptides from poliovirus type 1 capsid protein VP1 (negative control), ESAT-6 or Ag85A (see Methods for details). **B,** Proliferation of splenocytes from BCG::RD1-2F9-immunised mice in the absence or presence of ESAT-6 1-20 peptide, with or without anti-CD4 or anti-CD8 monoclonal antibody. Results are

expressed as mean and standard deviation of ^{3}H-thymidine incorporation from duplicate wells. **c,** Concentration of IFN-γ in culture supernatants of splenocytes of C57BL/6 mice stimulated for 72 h with peptides or PPD after s.c. or i.v. immunisation with either BCG::pYUB412 or BCG::RD1-2F9. Mice were inoculated with either 10^6 or 10^7 CFU. Results are expressed as the mean and standard deviation of duplicate culture wells.

Figure 14: Mouse protection studies. **A,** Bacterial counts in the spleen and lungs of vaccinated and unvaccinated C57BL/6 mice 2 months after an i.v. challenge with *M. tuberculosis* H37Rv. The mice were challenged 2 months after i.v. inoculation with 10^6 cfu of either BCG::pYUB412 or BCG::RD1-2F9. Organ homogenates for bacterial enumeration were plated on 7H11 medium, with or without hygromycin, to differentiate *M. tuberculosis* from residual BCG colonies. Results are expressed as the mean and standard deviation of 4 mice. Hatched columns correspond to the cohort of unvaccinated mice, while white and black columns correspond to mice vaccinated with BCG::pYUB412 and BCG::RD1-2F9, respectively. **B,** Bacterial counts in the spleen and lungs of vaccinated and unvaccinated C57BL6 mice after an aerosol challenge with 1000 CFUs of *M. tuberculosis*. All mice were treated with antibiotics for three weeks prior to infection with *M. tuberculosis*. Data are the mean and SE measured on groups of three animals, and differences between groups were analysed using ANOVA (*$p<0.05$, **$p<0.01$).

Figure 15: Guinea pig protection studies. **A,** Mean weight gain of vaccinated and unvaccinated guinea pigs following aerosol infection with *M. tuberculosis* H37Rv. Guinea pigs were vaccinated with either saline (triangles), BCG (squares) or BCG::RD1-2F9 (filled circles). The error bars are the standard error of the mean. Each time point represents the mean weight of six guinea pigs. For the saline vaccinated group the last live weight was used for calculating the means as the animals were killed on signs of severe tuberculosis which occurred after 50, 59, 71, 72, 93 and 93 days. **B,** Mean bacterial counts in the spleen and lungs of vaccinated and unvaccinated guinea pigs after

an aerosol challenge with *M. tuberculosis* H37Rv. Groups of 6 guinea pigs were vaccinated subcutaneously with either saline, BCG or BCG::RD1-2F9 and infected 56 days later. Vaccinated animals were killed 120 days following infection and unvaccinated ones on signs of suffering or significant weight loss. The error bars represent the standard error of the mean of six observations. **C,** Spleens of vaccinated guinea pigs 120 days after infection with *M. tuberculosis* H37Rv; left, animal immunised with BCG; right, animal immunised with BCG::RD1-2F9.

Figure 16: Diagram of the *M.tuberculosis* H37Rv genomic region showing a working model for biogenesis and export of ESAT-6 proteins. It presents a possible functional model indicating predicted subcellular localization and potential interactions within the mycobacterial cell envelope. Rosetta stone analysis indicates direct interaction between proteins Rv3870 and Rv3871,and the sequence similarity between the N-terminal domains of Rv3868 and Rv3876 suggests that these putative chaperones might also interact. Rv3868 is a member of the AAA-family of ATPases that perform chaperone-like functions by assisting in the assembly, and disassembly of protein complexes (Neuwald, A.F., Aravind, L., Spouge, J.L. & Koonin, E.V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**, 27-43. (1999).). It is striking that many type III secretion systems require chaperones for stabilisation of the effector proteins that they secrete and for prevention of premature protein-protein interactions (Page, A.L. & Parsot, C. Chaperones of the type III secretion pathway: jacks of all trades. *Mol Microbiol* **46**, 1-11. (2002).). Thus, Rv3868, and possibly Rv3876,may be required for the folding and/or dimerisation of ESAT-6/CFP-10 proteins (Renshaw, P.S., *et al.* Conclusive evidence that the major T-cell antigens of the *M. tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterisation of the structural properties of ESAT-6, CFP-10 and the ESAT-6-CFP-10 complex: implications for pathogenesis and virulence. *J Biol Chem* **8**, 8 (2002).),or even to prevent premature dimerisation. ESAT-6/CFP-10 are

predicted to be exported through a transmembrane channel,consisting of at least Rv3870, Rv3871, and Rv3877,and possibly Rv3869, in a process catalysed by ATP-hydrolysis. Rv3873 (PPE 68) is known to occur in the cell envelope and may also be involved as shown herein.

<u>Example 1: preparation and assessment of *M. bovis* BCG::RD1 strains as a vaccine for treating or preventing tuberculosis.</u>

As mentioned above, we have found that complementation with RD1 was accompanied by a change in colonial appearance as the BCG Pasteur "knock-in" strains developed a strikingly different morphotype (Fig. 2A). The RD1 complemented strains adopted a spreading, less-rugose morphology, that is characteristic of *M. bovis,* and this was more apparent when the colonies were inspected by light microscopy (Fig. 2B). Maps of the clones used are shown (Fig. 1C). These changes were seen following complementation with all of the RD1 constructs (Fig. 1C) and on complementing *M. microti* (data not shown). Pertinently, Calmette and Guérin (A. Calmette, *La vaccination preventive contre la tuberculose.* (Masson et cie., Paris, 1927)) observed a change in colony morphology during their initial passaging of *M. bovis*, and our experiments now demonstrate that this change, corresponding to loss of RD1, directly contributed to attenuating this virulent strain. The integrity of the cell wall is known to be a key virulence determinant for *M. tuberculosis* (C. E. Barry, *Trends Microbiol* 9, 237 (2001)), and changes in both cell wall lipids (M. S. Glickman, J. S. Cox, W. R. Jacobs, Jr., *Mol Cell* 5, 717 (2000)) and protein (F. X. Berthet, *et al.*, *Science* 282, 759 (1998)) have been shown to alter colony morphology and diminish persistence in animal models.

To determine which genes were implicated in these morphological changes, antibodies recognising three RD1 proteins (Rv3873, CFP10 and ESAT-6) were used in

immunocytological and subcellular fractionation analysis. When the different cell fractions from *M. tuberculosis* were immunoblotted all three proteins were localized in the cell wall fraction (Fig. 2C) though significant quantities of Rv3873, a PPE protein, were also detected in the membrane and cytosolic fractions (Fig. 2D). Using immunogold staining and electron microscopy the presence of ESAT-6 in the envelope of *M. tuberculosis* was confirmed but no alteration in capsular ultrastructure could be detected (data not shown). Previously, CFP-10 and ESAT-6 have been considered as secreted proteins (F. X. Berthet et al, *Microbiology* **144**, 3195 (1998)) but our results suggest that their biological functions are linked directly with the cell wall.

Changes in colonial morphology are often accompanied by altered bacterial virulence. Initial assessment of the growth of different BCG::RD1 "knock-ins" in C57BL/6 or BALB/c mice following intravenous infection revealed that complementation did not restore levels of virulence to those of the reference strain *M. tuberculosis* H37Rv (Fig. 3A). In longer-term experiments, modest yet significant differences were detected in the persistence of the BCG::RD1 "knock-ins" in comparison to BCG controls. Following intravenous infection of C57BL/6 mice, only the RD1 "knock-ins" were still detectable in the lungs after 106 days (Fig. 3B). This difference in virulence between the RD1 recombinants and the BCG vector control was more pronounced in severe combined immunodeficiency (SCID) mice (Fig. 3C). The BCG::RD1-2F9 "knock-in" was markedly more virulent, as evidenced by the growth rate in lungs and spleen and also by an increased degree of splenomegaly (Fig. 3D). Cytological examination revealed numerous bacilli, extensive cellular infiltration and granuloma formation. These increases in virulence following complementation with the RD1 region, demonstrate that the loss of this genomic locus contributed to the attenuation of BCG.

The inability to restore full virulence to BCG Pasteur was not due to instability of our constructs nor to the strain used (data not shown). Essentially identical results were obtained on complementing BCG Russia, a strain less passaged than BCG Pasteur and

presumed, therefore, to be closer to the original ancestor (M. A. Behr, *et al.*, *Science* **284**, 1520 (1999)). This indicates that the attenuation of BCG was a polymutational process and loss of residual virulence for animals was documented in the late 1920s (T. Oettinger, et al, *Tuber Lung Dis* **79**, 243 (1999)). Using the same experimental strategy, we also tested the effects of complementing with RD3-5, RD7 and RD9 (S. T. Cole, *et al., Nature 393, 537 (1998)* ; M. A. Behr, *et al., Science 284, 1520 (1999)* ; R. Brosch, *et al., Infection Immun. 66, 2221 (1998) and* S. V. Gordon *et al., Molec Microbiol 32, 643 (1999))* encoding putative virulence factors (Fig. 1B). Reintroduction of these regions, which are not restricted to avirulent strains, did not affect virulence in immuno-competent mice. Although it is possible that deletion effects act synergistically it seems more plausible that other attenuating mechanisms are at play.

Since RD1 encodes at least two potent T-cell antigens (R. Colangelli, *et al.*, *Infect. Immun.* **68**, 990 (2000), M. Harboe, et al., *Infect. Immun.* **66**, 717 (1998) and R. L. V. SkjØt, *et al.*, *Infect. Immun.* **68**, 214 (2000)), we investigated whether its restoration induced immune responses to these antigens or even improved the protective capacity of BCG. Three weeks following either intravenous or subcutaneous inoculation with BCG::RD1 or BCG controls, we observed similar proliferation of splenocytes to an Ag85A (an antigenic BCG protein) peptide (K. Huygen, et al., *Infect. Immun.* **62**, 363 (1994)), but not against a control viral peptide (Fig. 4A). Moreover, BCG::RD1 generated powerful $CD4^{+}$ T-cell responses against the ESAT-6 peptide as shown by splenocyte proliferation (Fig. 4A, B) and strong IFN-γ production (Fig. 4C). In contrast, the BCG::pYUB412 control did not stimulate ESAT-6 specific T-cell responses thus indicating that these were mediated by the RD1 locus. ESAT-6 is, therefore, highly immunogenic in mice in the context of recombinant BCG.

When used as a subunit vaccine, ESAT-6 elicits T-cell responses and induces levels of protection weaker than but akin to those of BCG (L. Brandt et al, *Infect. Immun.* **68**, 791 (2000)). Challenge experiments were conducted to determine if induction of immune

responses to BCG::RD1-encoded antigens, such as ESAT-6, could improve protection against infection with *M. tuberculosis*. Groups of mice inoculated with either BCG::pYUB412 or BCG::RD1 were subsequently infected intravenously with *M. tuberculosis* H37Rv. These experiments showed that immunisation with the BCG::RD1 "knock-in" inhibited the growth of *M. tuberculosis* within both BALB/c (Fig. 4D) and C57BL/6 mice when compared to inoculation with BCG alone.

Although the increases in protection induced by BCG::RD1 and the BCG control are modest they demonstrate convincingly that genetic differences have developed between the live vaccine and the pathogen which have weakened the protective capacity of BCG. This study therefore defines the genetic basis of a compromise that has occurred, during the attenuation process, between loss of virulence and reduced protection (M. A. Behr, P. M. Small, *Nature* **389**, 133 (1997)). The strategy of reintroducing, or even overproducing (M. A. Horwitz et al, *Proc Natl Acad Sci U S A* **97**, 13853 (2000)), the missing immunodominant antigens of *M. tuberculosis* in BCG, could be combined with an immuno-neutral attenuating mutation to create a more efficacious tuberculosis vaccine.

Example 2: BAC based comparative genomics identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant.

We searched for any genetic differences between human and vole isolates that might explain their different degree of virulence and host preference and what makes the vole isolates harmless for humans. In this regard, comparative genomics methods were employed in connection with the present invention to identify major differences that may exist between the *M. microti* reference strain OV254 and the entirely sequenced strains of *M. tuberculosis* H37Rv (10) or *M. bovis* AF2122/97 (14). An ordered Bacterial

Artificial Chromosome (BAC) library of *M. microti* OV254 was constructed and individual BAC to BAC comparison of a minimal set of these clones with BAC clones from previously constructed libraries of *M. tuberculosis* H37Rv and *M. bovis* AF2122/97 was undertaken.

Ten regions were detected in *M. microti* that were different to the corresponding genomic regions in *M. tuberculosis* and *M. bovis*. To investigate if these regions were associated with the ability of *M. microti* strains to infect humans, their genetic organization was studied in 8 additional *M. microti* strains, including those isolated recently from patients with pulmonary tuberculosis. This analysis identified some regions that were specifically absent from all tested *M. microti* strains, but present in all other members of the *M. tuberculosis* complex and other regions that were only absent from vole isolates of *M. microti*.

2.1 MATERIALS AND METHODS

Bacterial strains and plasmids. *M. microti* OV254 which was originally isolated from voles in the UK in the 1930's was kindly supplied by MJ Colston (45). DNA from *M. microti* OV216 and OV183 were included in a set of strains used during a multicenter study (26). *M. microti* Myc 94-2272 was isolated in 1988 from the perfusion fluid of a 41-year-old dialysis patient (43) and was kindly provided by L. M. Parsons. *M. microti* 35782 was purchased from American Type Culture Collection (designation TMC 1608 (M.P. Prague)). *M. microti* B1 type llama, B2 type llama, B3 type mouse and B4 type mouse were obtained from the collection of the National Reference Center for Mycobacteria, Forschungszentrum Borstel, Germany. *M. bovis* strain AF2122/97, spoligotype 9 was responsible for a herd outbreak in Devon in the UK and has been isolated from lesions in both cattle and badgers. Typically, mycobacteria were grown on

7H9 Middlebrook liquid medium (Difco) containing 10% oleic-acid-dextrose-catalase (Difco), 0.2 % pyruvic acid and 0.05% Tween 80.

Library construction, preparation of BAC DNA and sequencing reactions. Preparation of agarose-embedded genomic DNA from *M. microti* strain OV254, *M. tuberculosis* H37Rv, *M. bovis* BCG was performed as described by Brosch et al. (5). The *M. microti* library was constructed by ligation of partially digested *Hind*III fragments (50-125 kb) into pBeloBAC11. From the 10,000 clones that were obtained, 2,000 were picked into 96 well plates and stored at –80°C. Plasmid preparations of recombinant clones for sequencing reactions were obtained by pooling eight copies of 96 well plates, with each well containing an overnight culture in 250 µl 2YT medium with 12.5 $\mu g.ml^{-1}$ chloramphenicol. After 5 min centrifugation at 3000 rpm, the bacterial pellets were resuspended in 25 µl of solution A (25 mM Tris, pH 8.0, 50 mM glucose and 10 mM EDTA), cells were lysed by adding 25 µl of buffer B (NaOH 0.2 M, SDS 0.2%). Then 20 µl of cold 3 M sodium acetate pH 4.8 were added and kept on ice for 30 min. After centrifugation at 3000 rpm for 30 min, the pooled supernatants (140 µl) were transferred to new plates. 130 µl of isopropanol were added, and after 30 min on ice, DNA was pelleted by centrifugation at 3500 rpm for 15 min. The supernatant was discarded and the pellet resuspended in 50 µl of a 10 µg/ml RNAse A solution (in Tris 10 mM pH 7.5 / EDTA 10 mM) and incubated at 64°C for 15 min. After precipitation (2.5 µl of sodium acetate 3 M pH 7 and 200 µl of absolute ethanol) pellets were rinsed with 200 µl of 70% ethanol, air dried and finally suspended in 20 µl of TE buffer.

End-sequencing reactions were performed with a *Taq* DyeDeoxy Terminator cycle sequencing kit (Applied Biosystems) using a mixture of 13 µl of DNA solution, 2 µl of Primer (2 µM) (SP6-BAC1, AGTTAGCTCACTCATTAGGCA (SEQ ID No 15), or T7-BAC1, GGATGTGCTGCAAGGCGATTA (SEQ ID No 16)), 2.5 µl of Big Dye and 2.5 µl of a 5X buffer (50 mM $MgCl_2$, 50 mM Tris). Thermal cycling was performed on a

PTC-100 amplifier (MJ Inc.) with an initial denaturation step of 60 s at 95°C, followed by 90 cycles of 15 s at 95°C, 15 s at 56°c, 4 min at 60°C. DNA was then precipitated with 80 µl of 76% ethanol and centrifuged at 3000 rpm for 30 min. After discarding the supernatant, DNA was finally rinsed with 80 µl of 70% ethanol and resuspended in appropriate buffers depending on the type of automated sequencer used (ABI 377 or ABI 3700). Sequence data were transferred to Digital workstations and edited using the TED software from the Staden package (37). Edited sequences were compared against the *M. tuberculosis* H37Rv database (http://genolist.pasteur.fr/TubercuList/), the *M. bovis* BLAST server (http://www.sanger.ac.uk/Projects/M_bovis/blast_server.shtml), and in-house databases to determine the relative positions of the *M. microti* OV254 BAC end-sequences.

Preparation of BAC DNA from recombinants and BAC digestion profile comparison. DNA for digestion was prepared as previously described (4). DNA (1 µg) was digested with *Hind*III (Boehringer) and restriction products separated by pulsed-field gel electrophoresis (PFGE) on a Biorad CHEF-DR III system using a 1% (w/v) agarose gel and a pulse of 3.5 s for 17 h at 6 $V.cm^{-1}$. Low-range PFGE markers (NEB) were used as size standards. Insert sizes were estimated after ethidium bromide staining and visualization with UV light. Different comparisons were made with overlapping clones from the *M. microti* OV254, *M. bovis* AF2122/97, and *M. tuberculosis* H37Rv pBeloBAC11 libraries.

PCR analysis to determine presence of genes in different *M. microti* strains. Reactions contained 5 µl of 10xPCR buffer (100 mM ß-mercaptoethanol, 600 mM Tris-HCl, pH 8.8, 20 mM $MgCl_2$, 170 mM $(NH_4)_2SO_4$, 20 mM nucleotide mix dNTP), 2.5 µl of each primer at 2 µM, 10 ng of template DNA, 10% DMSO and 0.5 unit of Taq polymerase in a final volume of 12.5 µl. Thermal cycling was performed on a PTC-100

amplifier (MJ Inc.) with an initial denaturation step of 90 s at 95°C, followed by 35 cycles of 45 s at 95°C, 1 min at 60°C and 2 min at 72°C.

RFLP analysis. In brief, agarose plugs of genomic DNA prepared as previously described (5) were digested with either *Ase*I, *Dra*I or *Xba*I (NEB), then electrophoresed on a 1% agarose gel, and finally transferred to Hybond-C extra nitrocellulose membranes (Amersham). Different probes were amplified by PCR from the *M. microti* strain OV254 or *M. tuberculosis* H37Rv using primers for :

esat-6 (esat-6F GTCACGTCCATTCATTCCCT (SEQ ID No 17);

esat-6R ATCCCAGTGACGTTGCCTT) (SEQ ID No 18),

the RD1^{mic} flanking region (4340, 209F GCAGTGCAAAGGTGCAGATA (SEQ ID No 19); 4354,701R GATTGAGACACTTGCCACGA (SEQ ID No 20)), or

plcA (plcA.int.F CAAGTTGGGTCTGGTCGAAT (SEQ ID No 21); plcA.int.R GCTACCCAAGGTCTCCTGGT (SEQ ID No 22)). Amplification products were radio-labeled by using the Stratagene Prime-It II kit (Stratagene). Hybridizations were performed at 65°C in a solution containing NaCl 0.8 M, EDTA pH 8, 5 mM, sodium phosphate 50 mM pH 8, 2% SDS, 1X Denhardt's reagent and 100 μg/ml salmon sperm DNA (Genaxis). Membranes were exposed to phosphorimager screens and images were digitalized by using a STORM phospho-imager.

DNA sequence accession numbers. The nucleotide sequences that flank MiD1, MiD2, MiD3 as well as the junction sequence of RD1^{mic} have been deposited in the EMBL database. Accession numbers are AJ345005, AJ345006, AJ315556 and AJ315557, respectively.

2.2 RESULTS

Establishment of a complete ordered BAC library of *M. microti* OV254. Electroporation of pBeloBAC11 containing partial *Hind*III digests of *M. microti* OV254

DNA into *Escherichia coli* DH10B yielded about 10,000 recombinant clones, from which 2,000 were isolated and stored in 96-well plates. Using the complete sequence of the *M. tuberculosis* H37Rv genome as a scaffold, end-sequencing of 384 randomly chosen *M. microti* BAC clones allowed us to select enough clones to cover almost all of the 4.4 Mb chromosome. A few rare clones that spanned regions that were not covered by this approach were identified by PCR screening of pools as previously described (4). This resulted in a minimal set of 50 BACs, covering over 99.9% of the *M. microti* OV254 genome, whose positions relative to *M. tuberculosis* H37Rv are shown in Figure 5. The insert size ranged between 50 and 125 kb, and the recombinant clones were stable. Compared with other BAC libraries from tubercle bacilli (4, 13) the *M. microti* OV254 BAC library contained clones that were generally larger than those obtained previously, which facilitated the comparative genomics approach, described below.

Identification of DNA deletions in *M. microti* OV254 relative to *M. tuberculosis* H37Rv by comparative genomics. The minimal overlapping set of 50 BAC clones, together with the availability of three other ordered BAC libraries from *M. tuberculosis* H37Rv, *M. bovis* BCG Pasteur 1173P2 (5, 13) and *M. bovis* AF2122/97 (14) allowed us to carry out direct BAC to BAC comparison of clones spanning the same genomic regions. Size differences of PFGE-separated *Hind*III restriction fragments from *M. microti* OV254 BACs, relative to restriction fragments from *M. bovis* and/or *M. tuberculosis* BAC clones, identified loci that differed among the tested strains. Size variations of at least 2 kb were easily detectable and 10 deleted regions, evenly distributed around the genome, and containing more than 60 open reading frames (ORFs), were identified. These regions represent over 60 kb that are missing from *M. microti* OV254 strain compared to *M. tuberculosis* H37Rv. First, it was found that phiRv2 (RD11), one of the two *M. tuberculosis* H37Rv prophages was present in *M. microti* OV254, whereas phiRv1, also referred to as RD3 (29) was absent. Second, it was found that *M. microti* lacks four of the genomic regions that were also absent from *M.*

bovis BCG. In fact, these four regions of difference named RD7, RD8, RD9 and RD10 are absent from all members of the *M. tuberculosis* complex with the exception of *M. tuberculosis* and *M. canettii*, and seem to have been lost from a common progenitor strain of *M. africanum*, *M. microti* and *M. bovis* (3). As such, our results obtained with individual BAC to BAC comparisons show that *M. microti* is part of this non-*M. tuberculosis* lineage of the tubercle bacilli, and this assumption was further confirmed by sequencing the junction regions of RD7 – RD10 in *M. microti* OV254. The sequences obtained were identical to those from *M. africanum*, *M. bovis* and *M. bovis* BCG strains. Apart from these four conserved regions of difference, and phiRv1 (RD3) *M. microti* OV254 did not show any other RDs with identical junction regions to *M. bovis* BCG Pasteur, which misses at least 17 RDs relative to *M. tuberculosis* H37Rv (1, 13, 35). However, five other regions missing from the genome of *M. microti* OV254 relative to *M. tuberculosis* H37Rv were identified ($RD1^{mic}$, $RD5^{mic}$, MiD1, MiD2, MiD3). Such regions are specific either for strain OV254 or for *M. microti* strains in general. Interestingly, two of these regions, $RD1^{mic}$, $RD5^{mic}$ partially overlap RDs from the *M. bovis* BCG.

Antigens ESAT-6 and CFP-10 are absent from *M. microti*. One of the most interesting findings of the BAC to BAC comparison was a novel deletion in a genomic region close to the origin of replication (figure 5). Detailed PCR and sequence analysis of this region in *M. microti* OV254 showed a segment of 14 kb to be missing (equivalent to *M. tuberculosis* H37Rv from 4340,4 to 4354,5 kb) that partly overlapped $RD1^{bcg}$ absent from *M. bovis* BCG. More precisely, ORFs Rv3864 and Rv3876 are truncated in *M. microti* OV254 and ORFs Rv3865 to Rv3875 are absent (figure 6). This observation is particularly interesting as previous comparative genomic analysis identified $RD1^{bcg}$ as the only RD region that is specifically absent from all BCG sub-strains but present in all other members of the *M. tuberculosis* complex (1, 4, 13, 29, 35). As shown in Figure 6, in *M. microti* OV254 the $RD1^{mic}$ deletion is responsible for the loss of a large portion of

the conserved ESAT-6 family core region (40) including the genes coding for the major T-cell antigens ESAT-6 and CFP-10 (2, 15). The fact that previous deletion screening protocols employed primer sequences that were designed for the right hand portion of the $RD1^{bcg}$ region (i.e. gene Rv3878) (6, 39) explains why the $RD1^{mic}$ deletion was not detected earlier by these investigations. Figure 6 shows that $RD1^{mic}$ does not affect genes Rv3877, Rv3878 and Rv3879 which are part of the $RD1^{bcg}$ deletion.

Deletion of phospholipase-C genes in *M. microti* OV254. $RD5^{mic}$, the other region absent from *M. microti* OV254, that partially overlapped an RD region from BCG, was revealed by comparison of BAC clone Mi18A5 with BAC Rv143 (figure 5). PCR analysis and sequencing of the junction region revealed that $RD5^{mic}$ was smaller than the RD5 deletion in BCG (Table 2 and 3 below).

TABLE 2

Description of the putative function of the deleted and truncated ORFs in *M. microti* OV254

Region	Start - End	overlapping ORF	Putative Function or family
RD 10	264,5-266,5	Rv0221-Rv0223	*echA1*
RD 3	1779,5-1788,5	Rv1573-Rv1586	bacteriophage proteins
RD 7	2207,5-2220,5	Rv1964-Rv1977	*yrbE3A-3B; mce3A-F;* unknown
RD 9	2330-2332	Rv2072-Rv2075	*cobL;* probable oxidoreductase; unknown
$RD5^{mic}$	2627,6-2633,4	Rv2348-Rv2352	*plc A-C*; member of PPE family
MiD1	3121,8-3126,6	Rv2816-Rv2819	IS*6110* transposase; unknown
MiD2	3554,0-3755,2	Rv3187-Rv3190	IS*6110* transposase; unknown
MiD3	3741,1-3755,7	Rv3345-Rv3349	members of the PE-PGRS and PPE families; insertion elements
RD8	4056,8-4062,7	Rv3617-Rv3618	*ephA*; *lpqG;* member of the PE-PGRS family

$RD1^{mic}$	4340,4-4354,5	Rv3864-Rv3876	member of the CBXX/CF QX family; member of the PE and PPE families; ESAT-6; CFP10; unknown

TABLE 3. Sequence at the junction of the deleted regions in *M. microti* OV254

Junction	Position	ORFs	Sequences at the junction	Flanking primers
$RD1^{mic}$ (SEQ ID No 23)	4340,421-4354,533	Rv3864-Rv3876	**CAAGACGAGGTTGTAAAACCTCGACG CAGGATCGGCGATGAAATGCCAGTCG GCGTCGCTGAGCGCGCGCTGCGC***CGA GTCCCATTTTGTCGCTGATTTGTTTGAACA GCGACGAACCGGTGTTGAAAATGTCGCCT GGGTCGGGGATTCCCT*	4340,209F (SEQ ID No 19) GCAGTGCAAAGGTGCAGATA 4354,701R (SEQ ID No 20) GATTGAGACACTTGCCACGA
$RD5^{mic}$ (SEQ ID No 26)	2627,831-2635,581	Rv2349-Rv2355	**CCTCGATGAACCACCTGACATGACCC CATCCTTTCCAAGAACTGGAGTCTCC GGACATGCCGGGGCGGTTCA***CTGCCC CAGGTGTCCTGGGTCGTTCCGTTGACCGT CGAGTCCGAACATCCGTCATTCCCGGTGG CAGTCGGTGCGGTGAC*	2627,370F (SEQ ID No 24) GAATGCCGACGTCATATCG 2633,692R (SEQ ID No 25) CGGCCACTGAGTTCGATTAT
MiD1 (SEQ ID No 29)	3121,880-3126,684	Rv2815c-Rv2818c	**CACCTGACATGACCCCATCCTTTCCA AGAACTGGAGTCTCCGGACATGCCGG GGCGGTTCAG***GGACATTCATGTCCATCTT CTGGCAGATCAGCAGATCGCTTGTTCTCAG TGCAGGTGAGTC*	3121,690F (SEQ ID No 27) CAGCCAACACCAAGTAGACG 3126,924R (SEQ ID No 28) TCTACCTGCAGTCGCTTGTG
MiD2 (SEQ ID No 32)	3554,066-3555,259	Rv3188-Rv3189	**GCTGCCTACTACGCTCAACGCCAGAG ACCAGCCGCCGGCTGAGGTCTCAGAT CAGAGAGTCTCCGGACTCACCGGGGC GGTTCA***TAAAGGCTTCGAGACCGGACGG GCTGTAGGTTCCTCAACTGTGTGGCGGAT GGTCTGAGCACTTAAC*	3553,880F (SEQ ID No 30) GTCCATCGAGGATGTCGAGT 3555,385R (SEQ ID No 31) CTAGGCCATTCCGTTGTCTG
MiD3 (SEQ ID No 35)	3741,139-3755,777	Rv3345c-Rv3349c	**TGGCGCCGGCACCTCCGTTGCCACCG TTGCCGCCGCTGGTGGGCGCGGTGCC GTTCGCCCCGGCCGAACCGTTCAGGG CCGGGTTC***GCCCTCAGCCGCTAAACACG CCGACCAAGATCAACGAGCTACCTGCCCG GTCAAGGTTGAAGAGCCCCCATATCAGCA AGGGCCCGGTGTCGGCG*	3740,950F (SEQ ID No 33) GGCGACGCCATTTCC 3755,988R (SEQ ID No 34) AACTGTCGGGCTTGCTCTT

In fact, *M. microti* OV254 lacks the genes *plcA, plcB, plcC* and one specific PPE-protein encoding gene (Rv2352). This was confirmed by the absence of a clear band on a

Southern blot of *Ase*I digested genomic DNA from *M. microti* OV254 hybridized with a *plcA* probe. However, the genes Rv2346c and Rv2347c, members of the *esat-6* family, and Rv2348c, that are missing from *M. bovis* and BCG strains (3) are still present in *M. microti* OV254. The presence of an IS*6110* element in this segment suggests that recombination between two IS*6110* elements could have been involved in the loss of RD5^{mic}, and this is supported by the finding that the remaining copy of IS*6110* does not show a 3 base-pair direct repeat in strain OV254 (Table 3).

Lack of MiD1 provides genomic clue for *M. microti* OV254 characteristic spoligotype. MiD1 encompasses the three ORFs Rv2816, Rv2817 and Rv2818 that encode putative proteins whose functions are yet unknown, and has occurred in the direct repeat region (DR), a polymorphic locus in the genomes of the tubercle bacilli that contains a cluster of direct repeats of 36 bp, separated by unique spacer sequences of 36 to 41 bp (17), (figure 7). The presence or absence of 43 unique spacer sequences that intercalate the DR sequences is the basis of spacer-oligo typing, a powerful typing method for strains from the *M. tuberculosis* complex (23). *M. microti* isolates exhibit a characteristic spoligotype with an unusually small DR cluster, due to the presence of only spacers 37 and 38 (43). In *M. microti* OV254, the absence of spacers 1 to 36, which are present in many other *M. tuberculosis* complex strains, appears to result from an IS*6110* mediated deletion of 636 bp of the DR region. Amplification and *Pvu*II restriction analysis of a 2.8 kb fragment obtained with primers located in the genes that flank the DR region (Rv2813c and Rv2819) showed that there is only one copy of IS*6110* remaining in this region (figure 7). This IS*6110* element is inserted into ORF Rv2819 at position 3,119,932 relative to the *M. tuberculosis* H37Rv genome. As for other IS*6110* elements that result from homologous recombination between two copies (7), no 3 base-pair direct repeat was found for this copy of IS*6110* in the DR region. Concerning the absence of spacers 39-43 (figure7), it was found that *M. microti* showed a slightly different organization of this locus than *M. bovis* strains, which also

characteristically lack spacers 39-43. In *M. microti* OV254 an extra spacer of 36 bp was found that was not present in *M. bovis* nor in *M. tuberculosis* H37Rv. The sequence of this specific spacer was identical to that of spacer 58 reported by van Embden and colleagues (42). In their study of the DR region in many strains from the *M. tuberculosis* complex this spacer was only found in *M. microti* strain NLA000016240 (AF189828) and in some ancestral *M. tuberculosis* strains (3, 42). Like MiD1, MiD2 most probably results from an IS*6110*-mediated deletion of two genes (Rv3188, Rv3189) that encode putative proteins whose function is unknown (Table 3 above and Table 4 below).

TABLE 4. Presence of the RD and MiD regions in different *M. microti* strains

HOST	VOLES				HUMAN				
Strain	*OV 254*	OV183	*OV216*	ATCC 35782	*Myc 94 -2272*	B3 type mouse	B4 type mouse	B1 type llama	B2 type llama
RD1^{mic}	absent	absent	absent	absent	absent	absent	absent	absent	absent
RD 3	absent	absent	absent	absent	absent	absent	absent	absent	absent
RD 7	absent	absent	absent	absent	absent	absent	absent	absent	absent
RD8	absent	absent	absent	absent	absent	absent	absent	absent	absent
RD 9	absent	absent	absent	absent	absent	absent	absent	absent	absent
RD 10	absent	absent	absent	absent	absent	absent	absent	absent	absent
MiD3	absent	ND	ND	absent	absent	absent	absent	absent	absent
MiD1	absent	ND	ND	present	partial	partial	partial	present	present
RD5^{mic}	absent	absent	absent	present	present	present	present	present	present
MiD2	absent	ND	ND	present	present	present	present	present	present

ND, not determined

Absence of some members of the PPE family in *M. microti*. MiD3 was identified by the absence of two *Hind*III sites in BAC Mi4B9 that exist at positions 3749 kb and 3754 kb in the *M. tuberculosis* H37Rv chromosome. By PCR and sequence analysis, it was determined that MiD3 corresponds to a 12 kb deletion that has truncated or removed five genes orthologous to Rv3345c-Rv3349c. Rv3347c encodes a protein of 3157 amino-acids that belongs to the PPE family and Rv3346c a conserved protein that is also present in *M. leprae*. The function of both these putative proteins is unknown while Rv3348 and Rv3349 are part of an insertion element (Table 2). At present, the consequences of the MiD3 deletions for the biology of *M. microti* remains entirely unknown.

Extra-DNA in *M. microti* OV254 relative to *M. tuberculosis* H37Rv. *M. microti* OV254 possesses the 6 regions RvD1 to RvD5 and TBD1 that are absent from the sequenced strain *M. tuberculosis* H37Rv, but which have been shown to be present in other members of the *M. tuberculosis* complex, like *M. canettii*, *M. africanum*, *M. bovis*, and *M. bovis* BCG (3, 7, 13). In *M. tuberculosis* H37Rv, four of these regions (RvD2-5) contain a copy of IS*6110* which is not flanked by a direct repeat, suggesting that recombination of two IS*6110* elements was involved in the deletion of the intervening genomic regions (7). In consequence, it seems plausible that these regions were deleted from the *M. tuberculosis* H37Rv genome rather than specifically acquired by *M. microti*. In addition, three other small insertions have also been found and they are due to the presence of an IS*6110* element in a different location than in *M. tuberculosis* H37Rv and *M. bovis* AF2122/97. Indeed, *Pvu*II RFLP analysis of *M. microti* OV254 reveals 13 IS*6110* elements (data not shown).

Genomic diversity of *M. microti* strains. In order to obtain a more global picture of the genetic organization of the taxon *M. microti* we evaluated the presence or absence of the variable regions found in strain OV254 in eight other *M. microti* strains. These strains

which were isolated from humans and voles have been designated as *M. microti* mainly on the basis of their specific spoligotype (26, 32, 43) and can be further divided into subgroups according to the host such as voles, llama and humans (Table 3). As stated in the introduction, *M. microti* is rarely found in humans unlike *M. tuberculosis*. So the availability of 9 strains from variable sources for genetic characterization is an exceptional resource. Among them was one strain (Myc 94-2272) from a severely immuno-compromised individual (43), and four strains were isolated from HIV-positive or HIV-negative humans with spoligotypes typical of llama and mouse isolates. For one strain, ATCC 35872 / M.P. Prague, we could not identify with certainty the original host from which the strain was isolated, nor if this strain corresponds to *M. microti* OV166, that was received by Dr. Sula from Dr. Wells and used thereafter for the vaccination program in Prague in the 1960's (38).

First, we were interested if these nine strains designated as *M. microti* on the basis of their spoligotypes also resembled each other by other molecular typing criteria. As RFLP of pulsed-field gel separated chromosomal DNA represents probably the most accurate molecular typing strategy for bacterial isolates, we determined the *Ase*I profiles of the available *M. microti* strains, and found that the profiles resembled each other closely but differed significantly from the macro-restriction patterns of *M. tuberculosis*, *M. bovis* and *M. bovis* BCG strains used as controls. However, as depicted in Figure 8A, the patterns were not identical to each other and each *M. microti* strain showed subtle differences, suggesting that they were not epidemiologically related. A similar observation was made with other rare cutting restriction enzymes, like *Dra*I or *Xba*I (data not shown).

Common and diverging features of *M. microti* strains. Two strategies were used to test for the presence or absence of variable regions in these strains for which we do not have ordered BAC libraries. First, PCRs using internal and flanking primers of the

variable regions were employed and amplification products of the junction regions were sequenced. Second, probes from the internal portion of variable regions absent from *M. microti* OV254 were obtained by amplification of *M. tuberculosis* H37Rv DNA using specific primers. Hybridization with these radio-labeled probes was carried out on blots from PFGE separated *Ase*I restriction digests of the *M. microti* strains. In addition, we confirmed the findings obtained by these two techniques by using a focused macro-array, containing some of the genes identified in variable regions of the tubercle bacilli to date (data not shown).

This led to the finding that the RD1^{mic} deletion is specific for all *M. microti* strains tested.

Indeed, none of the *M. microti* DNA-digests hybridized with the radio-labeled *esat-6* probe (Fig. 8B) but with the RD1^{mic} flanking region (Fig. 8C). In addition, PCR amplification using primers flanking the RD1^{mic} region (Table 2) yielded fragments of the same size for *M. microti* strains whereas no products were obtained for *M. tuberculosis, M. bovis* and *M. bovis* BCG strains (Fig. 9). Furthermore, the sequence of the junction region was found identical among the strains which confirms that the genomic organization of the RD1^{mic} locus was the same in all tested *M. microti* strains (Table 3). This clearly demonstrates that *M. microti* lacks the conserved ESAT-6 family core region stretching in other members of the *M. tuberculosis* complex from Rv3864 to Rv3876 and, as such, represents a taxon of naturally occurring ESAT-6 / CFP-10 deletion mutants.

Like RD1^{mic}, MiD3 was found to be absent from all nine *M. microti* strains tested and, therefore, appears to be a specific genetic marker that is restricted to *M. microti* strains (Table 3). However, PCR amplification showed that RD5^{mic} is absent only from the vole isolates OV254, OV216 and OV183, but present in the *M. microti* strains isolated from human and other origins (Table 3). This was confirmed by the presence of single bands

but of differing sizes on a Southern blot hybridized with a *plcA* probe for all *M. microti* tested strains except OV254 (Fig. 8D). Interestingly, the presence or absence of $RD5^{mic}$ correlated with the similarity of IS*6110* RFLP profiles. The profiles of the three *M. microti* strains isolated from voles in the UK differed considerably from the IS*6110* RFLP patterns of humans isolates (43). Taken together, these results underline the proposed involvement of IS*6110* mediated deletion of the RD5 region and further suggest that RD5 may be involved in the variable potential of *M. microti* strains to cause disease in humans. Similarly, it was found that MiD1 was missing only from the vole isolates OV254, OV216 and OV183, which display the same spoligotype (43), confirming the observations that MiD1 confers the particular spoligotype of a group of *M. microti* strains isolated from voles. In contrast, PCR analysis revealed that MiD1 is only partially deleted from strains B3 and B4 both characterized by the mouse spoligotype and the human isolate *M. microti* Myc 94-2272 (Table 3). For strain ATCC 35782 deletion of the MiD1 region was not observed. These findings correlate with the described spoligotypes of the different isolates, as strains that had intact or partially deleted MiD1 regions had more spacers present than the vole isolates that only showed spacers 37 and 38.

2.3 COMMENTS AND DISCUSSION

We have searched for major genomic variations, due to insertion-deletion events, between the vole pathogen, *M. microti,* and the human pathogen, *M. tuberculosis.* BAC based comparative genomics led to the identification of 10 regions absent from the genome of the vole bacillus *M. microti* OV254 and several insertions due to IS*6110*. Seven of these deletion regions were also absent from eight other *M. microti* strains, isolated from voles or humans, and they account for more than 60 kb of genomic DNA.

Of these regions, $RD1^{mic}$ is of particular interest, because absence of part of this region has been found to be restricted to the BCG vaccine strains to date. As *M. microti* was originally described as non pathogenic for humans, it is proposed here that RD1 genes is involved in the pathogenicity for humans. This is reinforced by the fact that $RD1^{bcg}$ (29) has lost putative ORFs belonging to the *esat-6* gene cluster including the genes encoding ESAT-6 and CFP-10 (Fig. 6) (40). Both polypeptides have been shown to act as potent stimulators of the immune system and are antigens recognized during the early stages of infection (8, 12, 20, 34). Moreover, the biological importance of this RD1 region for mycobacteria is underlined by the fact that it is also conserved in *M. leprae*, where genes ML0047-ML0056 show high similarities in their sequence and operon organization to the genes in the *esat-6* core region of the tubercle bacilli (11). In spite of the radical gene decay observed in *M. leprae* the *esat-6* operon apparently has kept its functionality in this organism.

However, the RD1 deletion may not be the only reason why the vole bacillus is attenuated for humans. Indeed, it remains unclear why certain *M. microti* strains included in the present study that show exactly the same $RD1^{mic}$ deletion as vole isolates, have been found as causative agents of human tuberculosis. As human *M. microti* cases are extremely rare, the most plausible explanation for this phenomenon would be that the infected people were particularly susceptible for mycobacterial infections in general. This could have been due to an immunodeficiency (32, 43) or to a rare genetic host predisposition such as interferon gamma- or IL-12 receptor modification (22).

In addition, the finding that human *M. microti* isolates differed from vole isolates by the presence of region $RD5^{mic}$ may also have an impact on the increased potential of human *M. microti* isolates to cause disease. Intriguingly, BCG and the vole bacillus lack overlapping portions of this chromosomal region that encompasses three (*plcA*, *plcB*, *plcC*) of the four genes encoding phospholipase C (PLC) in *M. tuberculosis*. PLC has

been recognized as an important virulence factor in numerous bacteria, including *Clostridium perfringens*, *Listeria monocytogenes* and *Pseudomonas aeruginosa*, where it plays a role in cell to cell spread of bacteria, intracellular survival, and cytolysis (36, 41). To date, the exact role of PLC for the tubercle bacilli remains unclear. *plcA* encodes the antigen mtp40 which has previously been shown to be absent from seven tested vole and hyrax isolates (28). Phospholipase C activity in *M. tuberculosis*, *M. microti* and *M. bovis*, but not in *M. bovis* BCG, has been reported (21, 47). However, PLC and sphingomyelinase activities have been found associated with the most virulent mycobacterial species (21). The levels of phospholipase C activity detected in *M. bovis* were much lower than those seen in *M. tuberculosis* consistent with the loss of *plcABC*. It is likely, that *plcD* is responsible for the residual phospholipase C activity in strains lacking RD5, such as *M. bovis* and *M. microti* OV254. Indeed, the *plcD* gene is located in region RvD2 which is present in some but not all tubercle bacilli (13, 18). Phospholipase encoding genes have been recognized as hotspots for integration of IS*6110* and it appears that the regions RD5 and RvD2 undergo independent deletion processes more frequently than any other genomic regions (44). Thus, the virulence of some *M. microti* strains may be due to a combination of functional phospholipase C encoding genes (7, 25, 26, 29).

Another intriguing detail revealed by this study is that among the deleted genes seven code for members of the PPE family of Gly-, Ala-, Asn-rich proteins. A closer look at the sequences of these genes showed that in some cases they were small proteins with unique sequences, like for example Rv3873, located in the $RD1^{mic}$ region, or Rv2352c and Rv2353c located in the $RD5^{mic}$ region. Others, like Rv3347c, located in the MiD3 region code for a much larger PPE protein (3157 aa). In this case a neighboring gene (Rv3345c), belonging to another multigene family, the PE-PGRS family, was partly affected by the MiD3 deletion. While the function of the PE/PPE proteins is currently unknown, their predicted abundance in the proteome of *M. tuberculosis* suggests that

they may play an important role in the life cycle of the tubercle bacilli. Indeed, recently some of them were shown to be involved in the pathogenicity of *M. tuberculosis* strains (9). Complementation of such genomic regions in *M. microti* OV254 should enable us to carry out proteomics and virulence studies in animals in order to understand the role of such ORFs in pathogenesis.

In conclusion, this study has shown that *M. microti*, a taxon originally named after its major host *Microtus agrestis*, the common vole, represents a relatively homogenous group of tubercle bacilli. Although all tested strains showed unique PFGE macro-restriction patterns that differed slightly among each other, deletions that were common to all *M. microti* isolates (**RD7-RD10, MiD3, RD1^{mic}**) have been identified. The conserved nature of these deletions suggests that these strains are derived from a common precursor that has lost these regions, and their loss may account for some of the observed common phenotypic properties of *M. microti*, like the very slow growth on solid media and the formation of tiny colonies. This finding is consistent with results from a recent study that showed that *M. microti* strains carry a particular mutation in the *gyrB* gene (31).

Of particular interest, some of these common features (e.g. the flanking regions of RD1^{mic}, or MiD3) could be exploited for an easy-to-perform PCR identification test, similar to the one proposed for a range of tubercle bacilli (33). This test enables unambiguous and rapid identification of *M. microti* isolates in order to obtain a better estimate of the overall rate of *M. microti* infections in humans and other mammalian species.

<u>Example 3: Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis</u>

3.1 Complementation of the RD1 locus of BCG Pasteur and *M. microti*

To construct a recombinant vaccine that secretes both ESAT-6 and CFP-10, we complemented BCG Pasteur for the RD1 region using genomic fragments spanning variable sections of the *esxBA* (or ESAT-6) locus from *M. tuberculosis* (Fig. 10). The RD1 deletion in BCG interrupts or removes nine CDS and affects all four transcriptional units: three are removed entirely while the fourth (Rv3867-Rv3871) is largely intact apart from the loss of 112 codons from the 3'-end of Rv3871 (Fig. 10). Transcriptome analysis of BCG, performed using cDNA probes obtained from early log phase cultures with oligonucleotide-based microarrays, was able to detect signals at least two fold greater than background for the probes corresponding to Rv3867 to 3871 inclusive, but not for the RD1-deleted genes Rv3872 to Rv3879. This suggests that the Rv3867-3871 transcriptional unit is still active in BCG which, like *M. bovis*, also has frameshifts in the neighbouring gene, Rv3881 (Fig. 10). The RD1^{mic} deletion of *M. microti* removes three transcriptional units completely with only gene Rv3877 remaining from the fourth. The *M. tuberculosis* clinical isolate MT56 has lost genes Rv3878-Rv3879 (Brosch, R., *et al.* A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* 99, 3684-9. (2002)) but still secretes ESAT-6 and CFP-10 (Fig. 10).

To test the hypothesis that a dedicated export machinery exists and to establish which genes were essential for creating an ESAT-6-CFP-10 secreting vaccine we assembled a series of integrating vectors carrying fragments spanning different portions of the RD1 *esx* gene cluster (Fig. 10). These integrating vectors stably insert into the *attB* site of the genome of tubercle bacilli. pAP34 was designed to carry only the antigenic core region encoding ESAT-6 and CFP-10, and the upstream PE and PPE genes, whereas RD1-I106 and RD1-pAP35 were selected to include the core region and either the downstream or upstream portion of the gene cluster, respectively. The fourth construct RD1-2F9

contains a ~ 32 kb segment from *M. tuberculosis* that stretches from Rv3861 to Rv3885 covering the entire RD1 gene cluster. We adopted this strategy of complementation with large genomic fragments to avoid polar effects that might be expected if a putative protein complex is only partially complemented *in trans*. In addition, a set of smaller expression constructs (pAP47, pAP48) was established in which individual genes are transcribed from a heat shock promoter (Fig. 10). Using appropriate antibodies all of these constructs were found to produce the corresponding proteins after transformation of BCG or *M. microti* (see below).

3.2 Several genes of the *esx* cluster are required for export of ESAT-6 and CFP-10

The four BCG::RD1 recombinants (BCG::RD1-pAP34, BCG::RD1-pAP35, BCG::RD1-2F9 and BCG::RD1-I106) (Fig. 11) were initially tested to ensure that ESAT-6 and CFP-10 were being appropriately expressed from the respective integrated constructs. Immunoblotting of whole cell protein extracts from mid-log phase cultures of the various BCG::RD1 recombinants using an ESAT-6 monoclonal antibody or polyclonal sera for CFP-10 and the PPE68 protein Rv3873 demonstrated that all three proteins were expressed from the four constructs at levels comparable to those of *M. tuberculosis* (Fig. 11). However, striking differences were seen when the supernatants from early log-phase cultures of each recombinant were screened by Western blot for the two antigens. Although low levels of ESAT-6 and CFP-10 could be detected in the concentrated supernatant protein fractions of BCG::RD1-pAP34, BCG::RD1-pAP35 and BCG::RD1-I106 it was only with the integrated construct encompassing the entire *esx* gene cluster (BCG::RD1-2F9) that the two antigens accumulated in significant amounts. The high concentrations of ESAT-6 and CFP-10 seen in the supernatant of the recombinant BCG::RD1-2F9 were not due to a non-specific increase in permeability, or loss of cell wall material, because when the same whole cell and supernatant protein fractions were immunoblotted with serum raised against Rv3873, this protein was only localized in the

cell wall of the various recombinants. As expected, when constructs were used containing *esxA* or *esxBA* alone, ESAT-6 did not accumulate in the culture supernatant (data not shown).

To assess the effect of the $RD1^{mic}$ deletion of *M. microti* on the export of ESAT-6 and CFP-10 and subsequent antigen handling, the experiments were replicated in this genomic background. As with BCG, ESAT-6 and CFP-10 were only exported into the supernatant fraction in significant amounts if expressed in conjunction with the entire *esx* cluster (Fig. 11). The combined findings demonstrate that complementation with *esxA* or *esxB* alone is insufficient to produce a recombinant vaccine that secretes these two antigens. Rather, secretion requires expression of genes located both upstream and downstream of the antigenic core region confirming our hypothesis [20] that the conserved *esx* gene cluster does indeed encode functions essential for the export of ESAT-6 and CFP-10.

3.3 Secretion of ESAT-6 is needed to induce antigen specific T-cell responses

Since the classical observation that inoculation with live, but not dead BCG, confers protection against tuberculosis in animal models it has been considered that secretion of antigens is critical for maximizing protective T-cell immunity. Using our panel of recombinant vaccines we were able to test if antigen secretion was indeed essential for eliciting ESAT-6 specific T-cell responses. Groups of C57/BL6 mice were inoculated subcutaneously with one of six recombinant vaccines (BCG-pAP47, BCG-pAP48, BCG::RD1-pAP34, BCG::RD1-pAP35, BCG::RD1-I106, BCG::RD1-2F9) or with BCG transformed with the empty vector pYUB412. Three weeks following vaccination, T-cell immune responses to the seven vaccines were assessed by comparing antigen-specific splenocyte proliferation and gamma interferon (IFN-γ) production (Fig. 12A). As anticipated all of the vaccines generated splenocyte proliferation and IFN-γ production in

response to PPD (partially purified protein derivative) but not against an unrelated MalE control peptide indicating successful vaccination in each case. However, only splenocytes from the mice inoculated with BCG::RD1-2F9 proliferated markedly in response to the immunodominant ESAT-6 peptide (Fig. 12A). Furthermore, IFN-□ was only detected in culture supernatants of splenocytes from mice immunized with BCG::RD1-2F9 following incubation with the ESAT-6 peptide (Fig 12B) or recombinant CFP-10 protein (data not shown). These data demonstrate that export of the antigens is essential for stimulating specific Th1-oriented T-cells.

Further characterization of the immune responses was carried out. Splenocytes from mice immunized with BCG::RD1-2F9 or control BCG both proliferated in response to the immunodominant antigen 85A peptide (Fig 13A). The strong splenocyte proliferation in the presence of ESAT-6 was abolished by an anti-CD4 monoclonal antibody but not by anti-CD8 indicating that the $CD4^+$ T-cell subset was involved (Fig. 13B). Interestingly, as judged by *in vitro* IFN-γ response to PPD and the ESAT peptide, subcutaneous immunization generated much stronger T-cell responses (Fig. 13C) compared to intravenous injection. After subcutaneous immunisation with BCG::RD1-2F9 strong ESAT-6 specific responses were also detected in inguinal lymph nodes (data not shown). These experiments demonstrated that the ESAT-6 T-cell immune responses to vaccination with BCG::RD1-2F9 were potent, reproducible and robust making this recombinant an excellent candidate for protection studies.

3.4 Protective efficacy of BCG::RD1-2F9 in immuno-competent mice

When used alone as a subunit or DNA vaccine, ESAT-6 induces levels of protection weaker than but akin to those of BCG (Brandt, L., Elhay, M., Rosenkrands, I., Lindblad, E.B. & Andersen, P. ESAT-6 subunit vaccination against *Mycobacterium tuberculosis*. *Infect. Immun.* **68**, 791-795 (2000)). Thus, it was of interest to determine if the presentation to the immune system of ESAT-6 and/or CFP-10 in the context of recombinant BCG, mimicking the

presentation of the antigens during natural infection, could increase the protective efficiency of BCG. The BCG::RD1-2F9 recombinant was therefore selected for testing as a vaccine, since it was the only ESAT-6 exporting BCG that elicited vigorous antigen specific T-cell immune responses. Groups of C57BL/6 mice were inoculated intravenously with either BCG::RD1-2F9 or BCG::pYUB412 and challenged intravenously after eight weeks with *M. tuberculosis* H37Rv. Growth of *M. tuberculosis* H37Rv in spleens and lungs of each vaccinated cohort was compared with that of unvaccinated controls two months after infection (Fig. 14A). This demonstrated that, compared to vaccination with BCG, the BCG::RD1-2F9 vaccine inhibited growth of *M. tuberculosis* H37Rv in the spleens by 0.4 log10 CFU and was of comparable efficacy at protecting the lungs.

To investigate this enhanced protective effect against tuberculosis further we repeated the challenge experiment using the aerosol route. In this experiment antibiotic treatment was employed to clear persisting BCG from mouse organs prior to infection with *M. tuberculosis*. Two months following vaccination C57BL/6 mice were treated with daily rifampicin/izoniazid for three weeks and then infected with 1000 CFU of *M. tuberculosis* H37Rv by the respiratory route. Mice were then sacrificed after 17, 35 and 63 days and bacterial enumeration carried out on the lungs and spleen. This demonstrated that, even following respiratory infection, vaccination with BCG::RD1-2F9 was superior to vaccination with the control strain of BCG (Fig. 14B). However, growth of *M. tuberculosis* was again only inhibited strongly in the mouse spleens.

<u>**Example 4: Protective efficacy of BCG::RD1-2F9 in guinea pigs**</u>

4.1 Animal models *M. tuberculosis* H37Rv and the different recombinant vaccines were prepared in the same manner as for the immunological assays. For the guinea pig

assays, groups of outbred female Dunkin-Hartley guinea pigs (David Hall, UK) were inoculated with 5 x 10^4 CFUs by the subcutaneous route. Aerosol challenge was performed 8 weeks after vaccination using a contained Henderson apparatus and an H37Rv (NCTC 7416) suspension in order to obtain an estimated retained inhaled dose of approximately 1000 CFU/lung (Williams, A., Davies, A., Marsh, P.D., Chambers, M.A. & Hewinson, R.G. Comparison of the protective efficacy of bacille calmette-Guerin vaccination against aerosol challenge with *Mycobacterium tuberculosis* and *Mycobacterium bovis*. *Clin Infect Dis* **30 Suppl 3**, S299-301. (2000)). Organs were homogenized and dilutions plated out on 7H11 agar, as for the mice experiments. Guinea pig experiments were carried out in the framework of the European Union TB vaccine development program.

4.2 Results Although experiments in mice convincingly demonstrated a superior protective efficacy of BCG::RD1 over BCG it was important to establish a similar effect in the guinea pig model of tuberculosis. Guinea pigs are exquisitely sensitive to tuberculosis, succumbing rapidly to low dose infection with *M. tuberculosis*, and develop a necrotic granulomatous pathology closer to that of human tuberculosis. Immunization of guinea pigs with BCG::RD1-2F9 was therefore compared to conventional BCG vaccination. Groups of six guinea pigs were inoculated subcutaneously with saline, BCG or BCG::RD1-2F9. Eight weeks following inoculation the three guinea pig cohorts were challenged with *M. tuberculosis* H37Rv via the aerosol route. Individual animals were weighed weekly and were killed 17 weeks after challenge or earlier if they developed signs of severe tuberculosis. Whereas all unvaccinated guinea pigs failed to thrive and were euthanised before the last time-point because of overwhelming disease, both the BCG- and recombinant BCG::RD1-2F9-vaccinated animals progressively gained weight and were clinically well when killed on termination of the experiment (Fig. 15A). This indicated that although the BCG::RD1-2F9 recombinant is more virulent in severely immunodeficient mice (Pym, A.S., Brodin, P.,

Brosch, R., Huerre, M. & Cole, S.T. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol. Microbiol.* **46**, 709-717 (2002)). there is no increased pathogenesis in the highly susceptible guinea pig model of tuberculosis. Moreover, when the bacterial loads in the spleens of the vaccinated animals were compared there was a greater than ten-fold reduction in the number of CFU recovered from the animals immunised with BCG::RD1-2F9 when compared to BCG (Fig. 15B). Interestingly, there was no significant difference between the number of CFU obtained from the lungs of the two vaccinated groups indicating that the organ-specific enhanced protection observed in mice vaccinated with BCG::RD1-2F9 was also seen with guinea pigs. This marked reduction of bacterial loads in the spleens of BCG::RD1-2F9 immunised animals was also reflected in the gross pathology. Visual examination of the spleens showed that tubercules were much larger and more numerous on the surface of the BCG-vaccinated guinea pigs (Fig. 15C). These results demonstrate that the recombinant vaccine BCG::RD1-2F9 conveys enhanced protection to an aerosol challenge with *M. tuberculosis* in two distinct animal models.

GENERAL CONCLUSION

Tuberculosis is still one of the leading infectious causes of death in the world despite a decade of improving delivery of treatment and control strategies (Dye, C., Scheele, S., Dolin, P., Pathania, V. & Raviglione, M.C. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *Jama* **282**, 677-86. (1999)). Reasons for the recalcitrance of this pandemic are multi-factorial but include the modest efficacy of the widely used vaccine, BCG. Two broad approaches can be distinguished for the development of improved tuberculosis vaccines (Baldwin, S.L., *et al.* Evaluation of new

vaccines in the mouse and guinea pig model of tuberculosis. *Infection & Immunity* **66**, 2951-9 (1998), Kaufmann, S.H. How can immunology contribute to the control of tuberculosis□*Nature Rev Immunol* **1**, 20-30. (2001) and Young, D.B. & Fruth, U. in *New Generation Vaccines* (eds. Levine, M., Woodrow, G., Kaper, J. & Cobon GS) 631-645 (Marcel Dekker, 1997)). These are the development of subunit vaccines based on purified protein antigens or new live vaccines that stimulate a broader range of immune responses. Although a growing list of individual or combination subunit vaccines, and hybrid proteins, have been tested none has yet proved superior to BCG in animal models (Baldwin, S.L., *et al.,* 1998). Similarly, new attenuated vaccines derived from virulent *M. tuberculosis* have yet to out-perform BCG (Jackson, M., *et al.* Persistence and protective efficacy of a *Mycobacterium tuberculosis* auxotroph vaccine. *Infect Immun* **67**, 2867-73. (1999) and Hondalus, M.K., *et al.* Attenuation of and protection induced by a leucine auxotroph of *Mycobacterium tuberculosis. Infect Immun* **68**, 2888-98. (2000)). Interestingly, the only vaccine that appears to surpass BCG is a BCG recombinant over expressing antigen 85A (Horwitz, M.A., Harth, G., Dillon, B.J. & Maslesa-Galic, S. Recombinant bacillus calmette-guerin (BCG) vaccines expressing the Mycobacterium tuberculosis 30-kDa major secretory protein induce greater protective immunity against tuberculosis than conventional BCG vaccines in a highly susceptible animal model. *Proc Natl Acad Sci U S A* **97**, 13853-8. (2000)). The basis for this vaccine was the notion that over-expression of an immunodominant T-cell antigen could quantitatively enhance the BCG-elicited immune response.

In frame with the invention, we were able to show that restoration of the RD1 locus did indeed improve the protective efficacy of BCG and defines a genetic modification that should be included in new recombinant BCG vaccines. Moreover, we were able to demonstrate two further findings that will be crucial for the development of a live vaccine against tuberculosis. First, we have identified the genetic basis of secretion for

the ESAT-6 family of immunodominant T-cell antigens, and second, we show that export of these antigens from the cytosol is essential for maximizing their antigenicity.

The extra-cellular proteins of *M. tuberculosis* have been extensively studied and shown to be a rich source of protective antigens (Sorensen, A.L., Nagai, S., Houen, G., Andersen, P. & Andersen, A.B. Purification and characterization of a low-molecular-mass T-cell antigen secreted by Mycobacterium tuberculosis. *Infect Immun* **63**, 1710-7 (1995), SkjØt, R.L.V., *et al.* Comparative evaluation of low-molecular-mass proteins from *Mycobacterium tuberculosis* identifies members of the ESAT-6 family as immunodominant T-cell antigens. *Infect. Immun.* **68**, 214-220 (2000), Horwitz, M.A., Lee, B.W., Dillon, B.J. & Harth, G. Protective immunity against tuberculosis induced by vaccination with major extracellular proteins of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **92**, 1530-4 (1995) and Boesen, H., Jensen, B.N., Wilcke, T. & Andersen, P. Human T-cell responses to secreted antigen fractions of *Mycobacterium tuberculosis*. *Infect Immun* **63**, 1491-7 (1995)). Despite this it remains a mystery how some of these proteins, that lack conventional secretion signals, are exported from the cytosol, a unique problem in *M. tuberculosis* given the impermeability and waxy nature of the mycobacterial cell envelope. Although two *secA* orthologues were identified in the genome sequence of *M. tuberculosis* (Cole, S.T., *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537-544 (1998)), no genes for obvious type I, II, or III protein secretion systems were detected, like those that mediate the virulence of many Gram-negative bacterial pathogens (Finlay, B.B. & Falkow, S. Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* **61**, 136-169 (1997)). This suggested that novel secretion systems might exist. An *in silico* analysis of the *M. tuberculosis* proteome identified a set of proteins and genes whose inferred functions, genomic organisation and strict association with the *esx* gene family suggested that they could constitute such a system (Tekaia, F., *et al.* Analysis of the proteome of *Mycobacterium tuberculosis in silico*. *Tubercle Lung*

Disease 79, 329-342 (1999)). Our results provide the first empirical evidence that this gene cluster is essential for the normal export of ESAT-6 and CFP-10.

The antigen genes, *esxBA,* lie at the centre of the conserved gene cluster. Bioinformatics and comparative genomics predicted that both the conserved upstream genes Rv3868-Rv3871, as well as the downstream genes Rv3876-Rv3877, would be required for secretion (Fig. 1) and strong experimental support for this prediction is provided here. Our experiments show that only when BCG or *M. microti* are complemented with the entire cluster is maximal export of ESAT-6 and CFP-10 obtained. This suggests that at least Rv3871 and either Rv3876 or Rv3877 are indeed essential for the normal secretion of ESAT-6 as these are the only conserved genes absent or disrupted in BCG which are not complemented by RD1-I106 or RD1-pAP35. These genes encode a large transmembrane protein with ATPase activity, an ATP-dependent chaperone and an integral membrane protein, functional predictions compatible with them being part of a multi-protein complex involved in the translocation of polypeptides. Amongst the proteins encoded by the *esx* cluster Rv3871 and Rv3877 are highly conserved, as orthologues have been identified in the more streamlined clusters found in other actinomycetes, further supporting their direct role in secretion (Gey Van Pittius, N.C., *et al.* The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G+C Gram-positive bacteria. *Genome Biol* 2, 44.1-44.18 (2001)). It has been shown recently that ESAT-6 and CFP-10 form a heterodimer *in vitro* (Renshaw, P.S., *et al.* Conclusive evidence that the major T-cell antigens of the *M. tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterisation of the structural properties of ESAT-6, CFP-10 and the ESAT-6-CFP-10 complex: implications for pathogenesis and virulence. *J Biol Chem* 8, 8 (2002)) but it is not known whether dimerisation precedes translocation across the cell membrane or occurs at a later stage *in vivo*. In either case, chaperone or protein clamp activity is likely to be required to assist dimer formation or to prevent premature complexes arising as is well documented for type III secretion

systems (Page, A.L. & Parsot, C. Chaperones of the type III secretion pathway: jacks of all trades. *Mol Microbiol* 46, 1-11. (2002)). These, and other questions concerning the precise roles of the individual components of the ESAT-6 secretory apparatus, can now be addressed experimentally using the tools developed here.

The second major finding of the invention is that the secretion of ESAT-6 (and probably CFP-10) is critical for inducing maximal T-cell responses although other RD1-encoded proteins may also contribute such as the PPE68 protein (Rv3873) which is located in the cell envelope. We show that even though whole cell expression levels of ESAT-6 are comparable amongst our vaccines (Fig. 2), only the vaccine strain exporting ESAT-6, via an intact secretory apparatus, elicits powerful T-cell responses. Surprisingly, even the recombinants RD1-pAP47 and RD1-pAP48, that overexpress ESAT-6 intracellularly, did not generate detectable ESAT-6 specific T-cell responses. Although antigen secretion has long been recognized as important for inducing immunity against *M. tuberculosis*, and is often used to explain why killed BCG offers no protection, this is one of the first formal demonstrations of its importance. BCG, like *M. tuberculosis* resides in the phagosome, where secreted antigens have ready access to the MHC class II antigen processing pathway, essential for inducing IFN-γ producing CD4 T-cells considered critical for protection against tuberculosis. Further understanding of the mechanism of ESAT-6 secretion could allow the development of BCG recombinants that deliver other antigens in the same way.

The main aim of the present invention was to qualitatively enhance the antigenicity of BCG. So, having assembled a recombinant vaccine that secreted the T-cell antigens ESAT-6 and CFP-10, and shown that it elicited powerful CD4 T-cell immunity against at least ESAT-6 and CFP-10, the next step was to rigorously test its efficacy in animal models of tuberculosis. In three distinct models, including two involving respiratory challenge, we were able to demonstrate that the ESAT-6-CFP-10 secreting recombinant

improved protection when compared to a BCG control, although this effect was restricted to the spleen. This is probably due to the fact that the enhanced immunity induced by the two additional antigens is insufficient to abort the primary infection but does significantly reduce the dissemination of bacteria from the lung. The lack of protection afforded to the lung, the portal of entry for *M. tuberculosis*, does not prevent BCG::RD1-2F9 from being a promising vaccine candidate. Primary tuberculosis occurs in the middle and lower lobes and is rarely symptomatic (Garay, S.M. in *Tuberculosis* (eds. Rom, W.N. & Garay, S.M.) 373-413 (Little, Brown and Company, Boston, 1996)). The bacteria need to reach the upper lobes, the commonest site of disease, by haematogenous spread. Therefore, a vaccine that inhibits dissemination of *M. tuberculosis* from the primary site of infection would probably have major impact on the outcome of tuberculosis.

Recombinant BCG vaccines have definite advantages over other vaccination strategies in that they are inexpensive, easy to produce and convenient to store. However, despite an unrivalled and enviable safety record concerns remain and BCG is currently not administered to individuals with HIV infection. As shown above, the recombinant BCG::RD1-2F9 grows more rapidly in Severe Combined Immunodeficient (SCID) mice, an extreme model of immunodeficiency, than its parental BCG strain. However, in both immunocompetent mice and guinea pigs we have not observed any increased pathology only a slight increase in persistence which may be beneficial, since the declining efficacy of BCG with serial passage has been attributed to an inadvertent increase in its attenuation (Behr, M.A. & Small, P.M. Has BCG attenuated to impotence□*Nature* 389, 133-4. (1997)).

Ultimately, the robust enhancement in protection we have observed with the reincorporation of the RD1 locus is a compelling reason to include this genetic modification in any recombinant BCG vaccine, even if this may require the need for a balancing attenuating mutation.

In summary, the data presented here show that, in addition to its increased persistence, BCG::RD1-2F9 induces specific T-cell memory and enhances immune responses to other endogenous Th1 antigens such as the mycoloyl transferase, antigen 85A.

REFERENCES

1. **Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small.** 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science **284**:1520-1523.

2. **Berthet, F.-X., P. B. Rasmusse, I. Rosenkrands, P. Andersen, and B. Gicquel.** 1998. A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). Microbiology **144**:3195-3203.

3. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. U S A **99**:3684-3689.

4. **Brosch, R., S. V. Gordon, A. Billault, T. Garnier, K. Eiglmeier, C. Soravito, B. G. Barrell, and S. T. Cole.** 1998. Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect. Immun. **66**:2221-2229.

5. **Brosch, R., S. V. Gordon, C. Buchrieser, A. S. Pym, T. Garnier, and S. T. Cole.** 2000. Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. Comp. Funct. Genom. (Yeast) **17**:111-123.

6. **Brosch, R., S. V. Gordon, A. Pym, K. Eiglmeier, T. Garnier, and S. T. Cole.** 2000. Comparative genomics of the mycobacteria. Int. J. Med. Microbiol. **290**:143-152.

7. **Brosch, R., W. J. Philipp, E. Stavropoulos, M. J. Colston, S. T. Cole, and S. V. Gordon.** 1999. Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. Infect. Immun. 67:5768-5774.

8. **Brusasca, P. N., R. Colangeli, K. P. Lyashchenko, X. Zhao, M. Vogelstein, J. S. Spencer, D. N. McMurray, and M. L. Gennaro.** 2001. Immunological characterization of antigens encoded by the RD1 region of the *Mycobacterium tuberculosis* genome. Scand. J. Immunol. 54:448-452.

9. **Camacho, L. R., D. Ensergueix, E. Perez, B. Gicquel, and C. Guilhot.** 1999. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol. Microbiol. 34:257-267.

10. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, Barry C E, III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLeah, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Soeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537-544.

11. **Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T.**

Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, and B. G. Barrell. 2001. Massive gene decay in the leprosy bacillus. Nature **409**:1007-1011.

12. **Elhay, M. J., T. Oettinger, and P. Andersen.** 1998. Delayed-type hypersensitivity responses to ESAT-6 and MPT64 from *Mycobacterium tuberculosis* in the guinea pig. Infect. Immun. **66**:3454-3456.

13. **Gordon, S. V., R. Brosch, A. Billault, T. Garnier, K. Eiglmeier, and S. T. Cole.** 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol. Microbiol. **32**:643-655.

14. **Gordon, S. V., K. Eiglmeier, T. Garnier, R. Brosch, J. Parkhill, B. Barrell, S. T. Cole, and R. G. Hewinson.** 2001. Genomics of *Mycobacterium bovis*. Tuberculosis (Edinb) **81**:157-163.

15. **Harboe, M., A. S. Malin, H. S. Dockrell, H. G. Wiker, G. Ulvund, A. Holm, M. C. Jorgensen, and P. Andersen.** 1998. B-cell epitopes and quantification of the ESAT-6 protein of Mycobacterium tuberculosis. Infect. Immun. **66**:717-723.

16. **Hart, P. D. a., and I. Sutherland.** 1977. BCG and vole bacillus vaccines in the prevention of tuberculosis in adolescence and early adult life. British Medical Journal **2**:293-295.

17. **Hermans, P. W. M., D. Van Soolingen, E. M. Bik, P. E. W. De Haas, J. W. Dale, and J. D. A. van Embden.** 1991. Insertion element IS*987* from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. Infect. Immun. 59:2695-2705.

18. **Ho, T. B., B. D. Robertson, G. M. Taylor, R. J. Shaw, and D. B. Young.** 2000. Comparison of *Mycobacterium tuberculosis* genomes reveals frequent deletions in a 20 kb variable region in clinical isolates. Comp. Funct. Genom. (Yeast) 17:272-282.

19. **Horstkotte, M. A., I. Sobottka, K. Schewe Carl, P. Schaefer, R. Laufs, S. Ruesch-Gerdes, and S. Niemann.** 2001. *Mycobacterium microti* llama-type infection presenting as pulmonary tuberculosis in a human immunodeficiency virus-positive patient. J. Clin. Microbiol. 39:406-407.

20. **Horwitz, M. A., B. W. Lee, B. J. Dillon, and G. Harth.** 1995. Protective immunity against tuberculosis induced by vaccination with major extracellular proteins of *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U S A 92:1530-1534.

21. **Johansen, K. A., R. E. Gill, and M. L. Vasin.** 1996. Biochemical and molecular analysis of phospholipase C and phospholipase D activity in mycobacteria. Infect. Immun. 64:3259-3266.

22. **Jouanguy, E., R. Doffinger, S. Dupuis, A. Pallier, F. Altare, and J. L. Casanova.** 1999. IL-12 and IFN-gamma in host defense against mycobacteria and salmonella in mice and men. Curr. Opin. Immunol. 11:346-351.

23. **Kamerbeek, J., L. Schouls, A. Kolk, M. Van Agterveld, D. Van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal, and J. Van Embden.** 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. 35:907-914.

24. **Kato-Maeda, M., J. T. Rhee, T. R. Gingeras, H. Salamon, J. Drenkow, N. Smittipat, and P. M. Small.** 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res. **11**:547-554.

25. **Kremer, K., van Soolingen, D., van Embden, J., Hughes, S., Inwald, J., and G. Hewinson.** 1998. *Mycobacterium microti*: more widespread than previously thought. J. Clin. Microbiol. 36:2793-2794.

26. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden.** 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. 37:2607-2618.

27. **Levy Frebault, V., and F. Portaels.** 1992. Proposed minimal standards for the genus *Mycobacterium* and for description of new slowly growing *Mycobacterium* spp. Int.. J. Syst. Bact. **42**:315-323.

28. **Liebana, E., A. Aranaz, B. Francis, and D. Cousins.** 1996. Assessment of genetic markers for species differentiation within the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **34**:933-938.

29. **Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover.** 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J. Bact. **178**:1274-1282.

30. **Manabe, Y. C., Scott, C. P., and W. R. Bishai.** 2002 Naturally attenuated, orally administered *Mycobacterium microti* as a tuberculosis vaccine is better than subcutaneous *Mycobacterium bovis* BCG. Infect Immun. **70**:1566-1570.

31. **Niemann, S., Harmsen, D., Rusch-Gerdes, S., and E. Richter.** 2000. Differentiation of clinical *Mycobacterium tuberculosis* complex isolates by *gyrB* DNA sequence polymorphism analysis. J. Clin. Microbiol. **38**:3231-3234.

32. **Niemann, S., E. Richter, H. Daluegge-Tamm, H. Schlesinger, D. Graupner, B. Koenigstein, G. Gurath, U. Greinert, and S. Ruesch-Gerdes.** 2000. Two cases of *Mycobacterium microti*-derived tuberculosis in HIV-negative immunocompetent patients. Emerg. Infect. Dis. **6**:539-542.

33. **Parsons, L. M., Brosch, R., Cole, S. T., Somoskovi, A., Loder, A., Britzel, G., van Soolingen, D., Hale, Y., Salfinger, M.** 2002. Rapid and easy-to-perform identification of *Mycobacterium tuberculosis* complex isolates using PCR-based genomic deletion analysis. J. Clin. Microbiol. **submitted** and disclosure of the European Patent Application N° 02 290 458.2 filed on February 25, 2002 (Institut Pasteur).

34. **Rosenkrands, I., P. B. Rasmussen, M. Carnio, S. Jacobsen, M. Theisen, and P. Andersen.** 1998. Identification and characterization of a 29-kilodalton protein from *Mycobacterium tuberculosis* culture filtrate recognized by mouse memory effector cells. Infect. Immun. **66**:2728-2735.

35. **Salamon, H., M. Kato-Maeda, P. M. Small, J. Drenkow, and T. R. Gingeras.** 2000. Detection of deleted genomic DNA using a semiautomated computational analysis of GeneChip data. Genome Res. **10**:2044-2054.

36. **Songer, J. G.,** 1997. Bacterial phospholipases and their role in virulence. Trends Microbiol. 5:156-161.

37. **Staden, R.** 1996. The Staden sequence analysis package. Mol. Biotechnol. 5:233-241.

38. **Sula, L., and I. Radkovsky.** 1976. Protective effects of *M. microti* vaccine against tuberculosis. J. Hyg. Epid. Microbiol. Immunol. **20**:1-6.

39. **Talbot, E. A., D. L. Williams, and R. Frothingham.** 1997. PCR identification of *Mycobacterium bovis* BCG. J. Clin. Microbiol. 35:566-569.

40. **Tekaia, F., S. V. Gordon, T. Garnier, R. Brosch, B. G. Barrell, and S. T. Cole.** 1999. Analysis of the proteome of *Mycobacterium tuberculosis in silico.* Tubercle & Lung Disease **79**:329-342.

41. **Titball, R. W.** 1998. Bacterial phospholipases. Soc. Appl. Bacteriol. Symp. Ser. **27**:127-137.

42. **van Embden, J. D., T. van Gorkom, K. Kremer, R. Jansen, B. A. van Der Zeijst, and L. M. Schouls.** 2000. Genetic variation and evolutionary origin of the direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J. Bacteriol. **182**:2393-2401.

43. **van Soolingen, D., A. G. M. Van Der Zanden, P. E. W. De Haas, G. T. Noordhoek, A. Kiers, N. A. Foudraine, F. Portaels, A. H. J. Kolk, K. Kremer, and J. D. A. Van Embden.** 1998. Diagnosis *of Mycobacterium microti* infections among humans by using novel genetic markers. J. Clin. Microbiol. **36**:1840-1845.

44. **Vera-Cabrera, L., M. A. Hernandez-Vera, O. Welsh, W. M. Johnson, and J. Castro-Garza.** 2001. Phospholipase region of *Mycobacterium tuberculosis* is a preferential locus for IS*6110* transposition. J. Clin. Microbiol. **39**:3499-3504.

45. **Wells, A. Q.** 1937. Tuberculosis in wild voles. Lancet 1221.

46. **Wells, A. Q.** 1946. The murine type of tubercle bacillus. Medical Research council special report series **259**:1-42.

47. **Wheeler, P. R., and C. Ratledge.** 1992. Control and location of acyl-hydrolysing phospholipase activity in pathogenic mycobacteria. J. Gen. Microbiol. **138**:825-830.